

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2004, 09:16:45 ; Search time 97 Seconds  
(without alignments)  
783.560 Million cell updates/sec

Title: US-09-852-100B-2  
Perfect score: 1439  
Sequence: 1 MHILKGSPNVIPRAHGQKNT.....TRLTRLSITNETFRKTQLYP 269

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result							
No.	Score	Query	Match	Length	DB	ID	Description
1	1439	100.0	269	2	AAW94291		Aaw94291 Human bet
2	1439	100.0	269	3	AAY70759		Aay70759 Human bet
3	1439	100.0	269	6	AAE33877		Aae33877 Human BBP
4	748	52.0	139	2	AAY12358		Aay12358 Human 5'
5	673.5	46.8	162	2	AAY36021		Aay36021 Extended
6	667.5	46.4	148	2	AAY12426		Aay12426 Human 5'
7	429	29.8	81	7	ADB91834		Adb91834 Human sec
8	338	23.5	178	4	ABB65236		Abb65236 Drosophil
9	293	20.4	100	5	AAU97631		Aau97631 RNA polym

10	201	14.0	221	6	ADA57043	Ada57043	Human	sec
11	201	14.0	221	6	ABO14063	Abo14063	Novel	hum
12	201	14.0	221	6	ABR47818	Abr47818	Human	sec
13	201	14.0	221	6	ABR00112	Abr00112	Human	gen
14	201	14.0	221	7	ADB91589	Adb91589	Human	sec
15	201	14.0	221	7	ADC74204	Adc74204	Human	sec
16	201	14.0	222	2	AAY27658	Aay27658	Human	sec
17	200	13.9	221	3	AAY70761	Aay70761	Human	bet
18	200	13.9	221	6	ABR48493	Abr48493	Human	Amy
19	200	13.9	247	3	AAB08641	Aab08641	A human	T
20	200	13.9	247	3	AAB08656	Aab08656	A human	T
21	200	13.9	247	3	AAB08658	Aab08658	A human	T
22	200	13.9	247	3	AAB08657	Aab08657	A human	T
23	196	13.6	230	3	AAB08645	Aab08645	DNA	encod
24	182	12.6	284	4	ABB59014	Abb59014	Drosophil	
25	169.5	11.8	80	7	ADB91833	Adb91833	Human	sec
26	152.5	10.6	225	4	ABB11574	Abb11574	Human	sec
27	138.5	9.6	214	3	AAY87351	Aay87351	Human	sig
28	136.5	9.5	214	3	AAY70760	Aay70760	Human	bet
29	136.5	9.5	224	3	AAB43573	Aab43573	Human	can
30	132	9.2	224	4	ABB65466	Abb65466	Drosophil	
31	121	8.4	171	4	AAM93475	Aam93475	Human	pol
32	121	8.4	171	4	AAM93578	Aam93578	Human	pol
33	121	8.4	171	4	AAG67818	Aag67818	Human	gua
34	121	8.4	171	4	AAG81266	Aag81266	Human	AFP
35	121	8.4	171	5	ABB90376	Abb90376	Human	pol
36	121	8.4	171	6	ADA56861	Ada56861	Human	sec
37	121	8.4	171	6	ADA40712	Ada40712	Human	sec
38	121	8.4	172	2	AAY45270	Aay45270	Human	sec
39	119	8.3	140	5	ABR40585	Abr40585	Human	sec
40	115.5	8.0	213	5	ABB72307	Abb72307	Rat	prote
41	95.5	6.6	944	6	AAE32044	Aae32044	Human	ves
42	90.5	6.3	341	4	AAM41757	Aam41757	Human	pol
43	88.5	6.2	224	4	ABG20306	Abg20306	Novel	hum
44	88	6.1	68	6	AAE33878	Aae33878	Human	BBP
45	88	6.1	764	2	AAW94360	Aaw94360	Human	Fac

# ALIGNMENTS

## RESULT 1

AAW94291

ID AAW94291 standard; protein; 269 AA.

XX

AC AAW94291;

XX

DT 27-APR-1999 (first entry)

XX

DE Human beta-amyloid peptide-binding protein (BBP).

XX

KW Beta-amyloid peptide binding protein; BBP; beta-amyloid protein; BAP;  
KW human; Alzheimer's disease.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Region 68. .269  
 FT /note= "specifically claimed fragment having beta-amyloid  
 FT peptide binding activity"  
 XX  
 PN WO9846636-A2.  
 XX  
 PD 22-OCT-1998.  
 XX  
 PF 14-APR-1998; 98WO-US007462.  
 XX  
 PR 16-APR-1997; 97US-0064583P.  
 XX  
 PA (AMHP ) AMERICAN HOME PROD CORP.  
 XX  
 PI Ozenberger BA, Kajkowski EM, Jacobsen JS, Bard JA, Walker SG;  
 XX  
 DR WPI; 1999-080736/07.  
 DR N-PSDB; AAX05735.  
 XX  
 PT Polynucleotide encoding beta-amyloid peptide binding protein - used to  
 PT identify inhibitors of beta-amyloid peptide for treating Alzheimer's  
 PT disease.  
 XX  
 PS Claim 7; Page 43-44; 59pp; English.  
 XX  
 CC The present sequence represents a beta-amyloid peptide binding protein  
 CC (BBP). The polynucleotide comprising the entire BBP nucleotide sequence  
 CC of clone BBP1-fl is deposited under the accession number ATCC 98617. The  
 CC polynucleotide comprising a fragment of BBP (nucleotides 202-807 of the  
 CC full length BBP) of clone pEK196 is deposited as ATCC 98399. Host cells  
 CC transformed with a vector comprising the BBP nucleic acid are used for  
 CC the recombinant production of the protein. The protein can be used in a  
 CC method for diagnosing a disease characterised by aberrant expression of  
 CC human beta-amyloid protein (BAP). The protein can also be used in a  
 CC method for screening for compounds which regulate expression of a BAP  
 CC binding protein. The proteins, antibodies and identified compounds can be  
 CC used in the treatment or prevention of Alzheimer's disease  
 XX  
 SQ Sequence 269 AA;

Query Match 100.0%; Score 1439; DB 2; Length 269;  
 Best Local Similarity 100.0%; Pred. No. 8e-141;  
 Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MHILKGSPNVIPRAHGQKNTRRDGTGLYPMRGPFKNLALLPFSLPLLGGGGSGSGEKVSV	60
Db	1	MHILKGSPNVIPRAHGQKNTRRDGTGLYPMRGPFKNLALLPFSLPLLGGGGSGSGEKVSV	60
Qy	61	SKMAAAWPSGPSAPEAVTARLVGVLFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYIC	120
Db	61	SKMAAAWPSGPSAPEAVTARLVGVLFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYIC	120
Qy	121	KDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNG	180
Db	121	KDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNG	180
Qy	181	YSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCTVGF CGIGSLIDFILISMQIVGPSDG	240

```

Db      181 YSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCTVGFCGIGSLIDFILISMQIVGPSDG 240
Qy      241 SSYIIDYYGTRLTRLSITNETFRKTQLYP 269
Db      241 SSYIIDYYGTRLTRLSITNETFRKTQLYP 269

```

RESULT 2

AAAY70759

ID AAY70759 standard; protein; 269 AA.

XX

AC AAY70759;

XX

DT 24-JUL-2000 (first entry)

XX

DE Human beta-amyloid peptide (BAP) binding protein, BBP1.

XX

KW Beta-amyloid peptide binding protein; BBP; BAP; tumour; suppressor;  
 KW G-protein coupled receptor; GPCR; integral membrane protein; antigen;  
 KW neuronal cell; nonhuman primate; NHP; G-protein signalling pathway;  
 KW apoptosis; immunogen; therapeutic; treatment; prevention; diagnostic.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Domain 177. .198

FT /label= Transmembrane\_domain\_1

FT Domain 199. .201

FT /label= DRF\_motif

FT /note= "Substitution of the Arg abrogates protection"

FT Domain 213. .238

FT /label= Transmembrane\_domain\_2

XX

PN WO200022125-A2.

XX

PD 20-APR-2000.

XX

PF 13-OCT-1999; 99WO-US021621.

XX

PR 13-OCT-1998; 98US-0104104P.

XX

PA (AMHP ) AMERICAN HOME PROD CORP.

XX

PI Ozenberger BA, Kajkowski EM, Lo CF;

XX

DR WPI; 2000-317982/27.

DR N-PSDB; AAZ52369.

XX

PT Novel G-protein-coupled receptor-like proteins and polynucleotides useful  
 PT for regulating apoptosis, comprises integral membrane protein traversing  
 PT the membrane twice.

XX

PS Example 1; Page 62-63; 68pp; English.

XX

CC The present sequence is the beta-amyloid peptide (BAP) binding protein-1  
 CC (BBP1). It is an integral membrane protein, that traverse the membrane



CC twice. It is related to G protein-coupled receptor (GPCR) protein  
 CC superfamily. It interacts with G-alpha proteins and regulates the  
 CC activity of G-protein signalling pathways. BBP genes are widely expressed  
 CC in neuronal cells of nonhuman primate (NHP) brain and overexpressed in  
 CC some tumours. It functions as a suppressor of apoptosis induction. BBP  
 CC proteins are used as immunogens to raise antibodies, useful as  
 CC therapeutics and as antigens in solid phase assays. They are also useful  
 CC as reagents to identify molecules which effect the interaction of BBP and  
 CC a cloned protein, that are useful in the treatment or prevention of  
 CC diseases associated with apoptosis. The polynucleotides are useful for  
 CC diagnostics. Note: In claim 5, the patent claims an amino acid sequence  
 CC from figure 2. However, figure 2 does not contain any sequence. It is  
 CC inferred from the disclosure that the figure 2 sequence refers to BBP1  
 CC protein, shown in this sequence

XX

SQ Sequence 269 AA;

Query Match 100.0%; Score 1439; DB 3; Length 269;  
 Best Local Similarity 100.0%; Pred. No. 8e-141;  
 Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MHILKGSPNVIPRAHGQKNTRRDGTGLYPMRGPFKNLALLPFSPLLLGGGGSGSGEKVSV	60
Db	1	MHILKGSPNVIPRAHGQKNTRRDGTGLYPMRGPFKNLALLPFSPLLLGGGGSGSGEKVSV	60
Qy	61	SKMAAAWPSGPSAPEAVTARLVGVLFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYIC	120
Db	61	SKMAAAWPSGPSAPEAVTARLVGVLFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYIC	120
Qy	121	KDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNG	180
Db	121	KDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNG	180
Qy	181	YSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCTVGFCGIGSLIDFILISMQIVGPSDG	240
Db	181	YSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCTVGFCGIGSLIDFILISMQIVGPSDG	240
Qy	241	SSYIIDYYGTRLTRLSITNETFRKTQLYP	269
Db	241	SSYIIDYYGTRLTRLSITNETFRKTQLYP	269

# RESULT 3

AAE33877

ID AAE33877 standard; protein; 269 AA.

XX

AC AAE33877;

XX

DT 02-MAY-2003 (first entry)

XX

DE Human BBP-1 protein.

XX

KW Human; beta-amyloid peptide-binding protein; BAP; Abeta; betaAP; BBP;

KW Alzheimer's disease; AD; transgenic; transgenic animal; gene therapy;

KW neuroprotective; nootropic.

XX

OS Homo sapiens.

XX  
 PN WO200290499-A2.  
 XX  
 PD 14-NOV-2002.  
 XX  
 PF 06-MAY-2002; 2002WO-US014223.  
 XX  
 PR 09-MAY-2001; 2001US-00852100.  
 XX  
 PA (AMHP ) WYETH.  
 XX  
 PI Ozenberger BA, Bard JA, Kajkowski EM, Jacobsen JS, Walker SG;  
 PI Sofia HJ, Howland DS;  
 XX  
 DR WPI; 2003-120537/11.  
 DR N-PSDB; AAD51940.  
 XX  
 PT New human beta-amyloid peptide-binding protein, useful for diagnosing  
 PT and/or treating diseases associated with aberrant expression of beta-  
 PT amyloid peptide, e.g. Alzheimer's disease.  
 XX  
 PS Claim 4; Page 84-85; 85pp; English.  
 XX  
 CC The present invention relates to novel human beta-amyloid peptide (BAP;  
 CC Abeta, betaAP)-binding (BBP) proteins and polynucleotides encoding such  
 CC proteins. BBP sequences are useful to diagnose and/or treat diseases  
 CC associated with aberrant expression of human BAP such as Alzheimer's  
 CC disease (AD). They are used to generate transgenic animals. Sequences of  
 CC the invention are also used in gene therapy. The present sequence is  
 CC human BBP-1 protein  
 XX  
 SQ Sequence 269 AA;

Query Match 100.0%; Score 1439; DB 6; Length 269;  
 Best Local Similarity 100.0%; Pred. No. 8e-141;  
 Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MHILKGSPNVIPRAHGQKNTRRDGTGLYPMRGPFKNLALLPFSLP	60
Db	1	MHILKGSPNVIPRAHGQKNTRRDGTGLYPMRGPFKNLALLPFSLP	60
Qy	61	SKMAAAWPSGPSAPEAVTARLVGVLFVSVTTGPWGAVATSAGGEESLKCEDLKV	120
Db	61	SKMAAAWPSGPSAPEAVTARLVGVLFVSVTTGPWGAVATSAGGEESLKCEDLKV	120
Qy	121	KDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPI	180
Db	121	KDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPI	180
Qy	181	YSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCTVGFCGIGSLIDFILISM	240
Db	181	YSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCTVGFCGIGSLIDFILISM	240
Qy	241	SSYIIDYYGTRLTRLSITNETFRKTQLYP	269
Db	241	SSYIIDYYGTRLTRLSITNETFRKTQLYP	269

RESULT 4

AAY12358

ID AAY12358 standard; protein; 139 AA.

XX

AC AAY12358;

XX

DT 17-JUN-1999 (first entry)

XX

DE Human 5' EST secreted protein SEQ ID NO:389.

XX

KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
KW forensic; gene therapy; chromosome mapping; signal peptide;  
KW upstream regulatory sequence; cytokine activity; cell proliferation;  
KW differentiation; haematopoiesis regulation; tissue growth regulation;  
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; anti-inflammatory; tumour inhibition.

XX

OS Homo sapiens.

XX

PN WO9906548-A2.

XX

PD 11-FEB-1999.

XX

PF 31-JUL-1998; 98WO-IB001222.

XX

PR 01-AUG-1997; 97US-00905135.

XX

PA (GEST ) GENSET.

XX

PI Dumas Milne Edwards J, Duclert A, Lacroix B;

XX

DR WPI; 1999-153778/13.

DR N-PSDB; AAX41191.

XX

PT New nucleic acids encoding human secreted proteins - obtained from cDNA  
PT libraries prepared from e.g. liver, ovary, brain, prostate, kidney, lung,  
PT umbilical cord, placenta and colon tissue.

XX

PS Claim 27; Page 714-715; 824pp; English.

XX

CC AAX41094 to AAX41347 represent 5' expressed sequence tags (ESTs) for  
CC human secreted proteins, and encode the proteins given in AAY12261 to  
CC AAY12514, respectively. The proteins given represent the signal peptide  
CC and an N-terminal fragment of a secreted protein. The nucleic acid  
CC sequences can be used for producing secreted human gene products. They  
CC can also be used to develop products for diagnosis and therapy. The  
CC proteins obtained may have cytokine activity, cell  
CC proliferation/differentiation activity, haematopoiesis regulating  
CC activity, tissue growth regulating activity, reproductive hormone  
CC regulating activity, chemotactic/ chemokinetic activity, haemostatic and  
CC thrombolytic activity, receptor/ ligand activity, anti-inflammatory  
CC activity, tumour inhibition activity or other activities. The products  
CC can be used in forensic, gene therapy and chromosome mapping procedures.  
CC The sequences can also be used for obtaining corresponding promoter  
CC sequences. The nucleic acids encoding the signal peptide can be used for  
CC directing extracellular secretion of a polypeptide or the insertion of a

CC polypeptide into a membrane, or importing a polypeptide into a cell  
XX  
SQ Sequence 139 AA;

Query Match 52.0%; Score 748; DB 2; Length 139;  
Best Local Similarity 99.3%; Pred. No. 2.2e-69;  
Matches 138; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 63 MAAAWPSGPSAPEAVTARLVGVLFVSVTTGPWGAVATSAGGEESLKCEDLVKGQYICKD 122  
||||| |||||||||||||||||||||||||||||||||||||||  
Db 1 MAAAWXSGPSAPEAVTARLVGVLFVSVTTGPWGAVATSAGGEESLKCEDLVKGQYICKD 60  
  
Qy 123 PKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYS 182  
||||| |||||||||||||||||||||||||||||||||||||||  
Db 61 PKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYS 120  
  
Qy 183 YKVAVALSLFLGWLGADRF 201  
|||||  
Db 121 YKVAVALSLFLGWLGADRF 139

RESULT 5

AAAY36021

ID AAY36021 standard; protein; 162 AA.

XX

AC AAY36021;

XX

DT 13-SEP-1999 (first entry)

XX

DE Extended human secreted protein sequence, SEQ ID NO. 406.

XX

KW Secreted protein; human; cytokine; cellular proliferation; cell movement;  
KW cellular differentiation; immune system regulator; anti-inflammatory;  
KW haematopoiesis regulator; tissue growth regulator; tumour inhibitor;  
KW reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;  
KW genetic disease.

XX

OS Homo sapiens.

XX

PN WO9931236-A2.

XX

PD 24-JUN-1999.

XX

PF 17-DEC-1998; 98WO-IB002122.

XX

PR 17-DEC-1997; 97US-0069957P.

PR 09-FEB-1998; 98US-0074121P.

PR 13-APR-1998; 98US-0081563P.

PR 10-AUG-1998; 98US-0096116P.

XX

PA (GEST ) GENSET.

XX

PI Bougueleret L, Duclert A, Dumas Milne Edwards J;

XX

DR WPI; 1999-385906/32.

DR N-PSDB; AAX97705.

XX

PT New isolated human secreted proteins.  
 XX  
 PS Claim 9; Page 346-347; 516pp; English.  
 XX  
 CC This sequence is encoded by an extended human secreted protein coding  
 CC sequence of the invention. The secreted proteins can be used in treating  
 CC or controlling a variety of human conditions. The secreted proteins may  
 CC act as cytokines or may affect cellular proliferation or differentiation  
 CC or may act as immune system regulators, haematopoiesis regulators, tissue  
 CC growth regulators, regulators of reproductive hormones or cell movement  
 CC or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or  
 CC tumour inhibition activity. The DNAs can be used in forensic procedures  
 CC to identify individuals or in diagnostic procedures to identify  
 CC individuals having genetic diseases resulting from abnormal expression of  
 CC the genes corresponding to the extended cDNAs. They are also useful for  
 CC constructing a high resolution map of the human chromosomes. They can  
 CC also be used for gene therapy to control or treat genetic diseases  
 XX  
 SQ Sequence 162 AA;

Query Match 46.8%; Score 673.5; DB 2; Length 162;  
 Best Local Similarity 84.1%; Pred. No. 1.6e-61;  
 Matches 127; Conservative 4; Mismatches 17; Indels 3; Gaps 2;

Qy 63 MAAAWPSGPSAPEAVTARLVGVLFVSVTTGPWGAVATSAGGEESLKCEDLVKGQYICKD 122  
 |||||  
 Db 1 MAAAWPSGPXAPEAVTARLVGVLFVSVTTGPWGAVATSAGGEESLKCEDLVKGQYICKD 60  
 Qy 123 PKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYS 182  
 |||||  
 Db 61 PKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYS 120  
 Qy 183 YKVAVALSLFLGWLGADRFYLGYPALGLLK 212  
 | : | : | : | : |  
 Db 121 YNEQSHVS--FSWMVGSR SILPWIPCFGFVK 149

# RESULT 6

AAY12426

ID AAY12426 standard; protein; 148 AA.

XX

AC AAY12426;

XX

DT 17-JUN-1999 (first entry)

XX

DE Human 5' EST secreted protein SEQ ID NO:457.

XX

KW Human; secreted protein; EST; expressed sequence tag; diagnosis;

KW forensic; gene therapy; chromosome mapping; signal peptide;

KW upstream regulatory sequence; cytokine activity; cell proliferation;

KW differentiation; haematopoiesis regulation; tissue growth regulation;

KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;

KW thrombolytic; anti-inflammatory; tumour inhibition.

XX

OS Homo sapiens.

XX

PN WO9906548-A2.

```

XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-IB001222.
XX
PR 01-AUG-1997; 97US-00905135.
XX
PA (GEST ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Lacroix B;
XX
DR WPI; 1999-153778/13.
DR N-PSDB; AAX41259.
XX
PT New nucleic acids encoding human secreted proteins - obtained from cDNA
PT libraries prepared from e.g. liver, ovary, brain, prostate, kidney, lung,
PT umbilical cord, placenta and colon tissue.
XX
PS Claim 27; Page 763-764; 824pp; English.
XX
CC AAX41094 to AAX41347 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins, and encode the proteins given in AAY12261 to
CC AAY12514, respectively. The proteins given represent the signal peptide
CC and an N-terminal fragment of a secreted protein. The nucleic acid
CC sequences can be used for producing secreted human gene products. They
CC can also be used to develop products for diagnosis and therapy. The
CC proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell
XX
SQ Sequence 148 AA;

```

```

Qy      63 MAAAWPSGPSAPEAVTARLVGVLWVSVTTGPWGAVATSAGGEESLKCEDLKVGQYICKD 122
      |||
Db      1 MAAAWPSGPXAPEAVTARLVGVLWVSVTTGPWGAVATSAGGEESLKCEDLKVGQYICKD 60

Qy     123 PKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYS 182
      |||
Db      61 PKINDATQEPVNCTNYTAHVSCFPAPNITCKDXSGNETHFTGNEVGFFKPISCRNVNGYS 120

Qy     183 YKVAVALSLFLGWLGADR FYLG Y-PALGLL 211
      | : | : | : | :
Db     121 YXXQXXVS--FSWMVGSR SILPWIPC GFV 148

```

RESULT 7

ADB91834

ID ADB91834 standard; protein; 81 AA.

XX

AC ADB91834;

XX

DT 04-DEC-2003 (first entry)

XX

DE Human secreted protein #SEQ ID 780.

XX

KW Secreted protein; gene therapy; antidiabetic; diabetes; human.

XX

OS Homo sapiens.

XX

PN WO2003004622-A2.

XX

PD 16-JAN-2003.

XX

PF 19-MAR-2002; 2002WO-US008124.

XX

PR 21-MAR-2001; 2001US-0277340P.

PR 19-JUL-2001; 2001US-0306171P.

PR 13-NOV-2001; 2001US-0331287P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Rosen CA, Ruben SM;

XX

DR WPI; 2003-229407/22.

XX

PT Nucleic acid encoding a human secreted protein is useful in diagnosing or  
PT treating diabetes or conditions related to diabetes.

XX

PS Claim 3; SEQ ID NO 780; 1537pp; English.

XX

CC The invention relates to isolated nucleic acid molecules ADB91065-  
CC ADB91448 and ADB91835-ADB91911 encoding human secreted proteins ADB91449-  
CC ADB91834. Also disclosed is a recombinant vector comprising a  
CC polynucleotide of the invention, and a recombinant host cell comprising  
CC the recombinant vector. The polypeptide of the invention is useful in  
CC identifying a binding partner by contacting the polypeptide with a  
CC binding partner, and determining whether the binding partner increases or  
CC decreases activity of the polypeptide. The polypeptide, polynucleotide,  
CC antibody or its fragment, agonist or antagonist are useful for preparing  
CC a pharmaceutical composition for diagnosing or treating diabetes or  
CC conditions related to diabetes. The present sequence is that of the human  
CC immunoglobulin Fc portion used to generate fusion proteins, increasing  
CC the stability of the fused protein as compared to the secreted protein  
CC only. Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 81 AA;

Query Match 29.8%; Score 429; DB 7; Length 81;

Best Local Similarity 100.0%; Pred. No. 1.5e-36;

Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 164 GNEVGFFKPISCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCTVGFCGIGS 223  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 GNEVGFFKPISCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCTVGFCGIGS 60  
 Qy 224 LIDFILISMQIVGPSDGSSYI 244  
 ||||||||||||||||  
 Db 61 LIDFILISMQIVGPSDGSSYI 81

RESULT 8

ABB65236

ID ABB65236 standard; protein; 178 AA.

XX

AC ABB65236;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 22500.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US009231.

XX

PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX

PA (PEKE ) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656860/75.

DR N-PSDB; ABL09339.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signaling and cell-cell

PT interactions.

XX

PS Disclosure; SEQ ID NO 22500; 21pp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences



XX  
SQ Sequence 178 AA;

Query Match 23.5%; Score 338; DB 4; Length 178;  
Best Local Similarity 42.6%; Pred. No. 1.4e-26;  
Matches 69; Conservative 30; Mismatches 49; Indels 14; Gaps 5;

```
Qy      107 SLKCEDLK-VGQYICKDP---KINDATQEPVNCTNY-TAHVSCFPAPNITCKDSSGNETH 161
          :: | :: :||:| ||  ::: ||:  ||  | | | | | | : :|| |
Db      20 NVDCNELQMMGQFMCPDPARGQIDPKTQQLAGCTREGRARVWCIAANEINCTE-TGNAT- 77

Qy      162 FTGNEVGFFKPISCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCFTVGFCGI 221
          | : : |: |||  : ||:| | | ||||| | :||| | | :| :
Db      78 -----FTREVPCKWTNGYHLDTTLLLSVFLGMFGVDRFYLGYPGIGLLKFCTLGGMFL 130

Qy      222 GSLIDFILISMQIVGPSDGSSYIIDYYGTRLRLSITNETFR 263
          | ||| :||:|:|:|:|:|:|:| | | | : : | | :|
Db      131 GQLIDIVLIALQVVGPADGSAYVIPYYGAGIHIVRSNTTYR 172
```

RESULT 9

AAU97631

ID AAU97631 standard; protein; 100 AA.

XX

AC AAU97631;

XX

DT 13-AUG-2002 (first entry)

XX

DE RNA polymerase II subunit 11 protein.

XX

KW RNA polymerase II subunit 11; cancer; HIV; infection;

KW human immunodeficiency virus.

XX

OS Unidentified.

XX

PN CN1331300-A.

XX

PD 16-JAN-2002.

XX

PF 30-JUN-2000; 2000CN-00116963.

XX

PR 30-JUN-2000; 2000CN-00116963.

XX

PA (BODE-) BODE GENE DEV CO LTD SHANGHAI.

XX

PI Mao Y, Xie Y;

XX

DR WPI; 2002-340664/38.

DR N-PSDB; ABK52558.

XX

PT Polypeptide-RNA polymerase II subunit 11 and polynucleotide for coding  
PT it.

XX

PS Claim 1; Page 29; 32pp; Chinese.

XX

CC This invention relates to the DNA and protein sequences of a novel  
CC polypeptide-RNA polymerase II subunit 11 protein. The invention also

CC comprises a process for preparing the polypeptide of the invention by DNA  
CC recombination, the application of the polypeptide in treating diseases  
CC such as cancer, human immunodeficiency virus (HIV) infection, etc, the  
CC antagonist of the polypeptide and its medical action, and the application  
CC of the said polynucleotide are disclosed. The present sequence represents  
CC the RNA polymerase II subunit 11 protein of the invention

XX

SQ Sequence 100 AA;

Query Match 20.4%; Score 293; DB 5; Length 100;

Best Local Similarity 98.2%; Pred. No. 2.9e-22;

Matches 55; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 63 MAAAWPSGPSAPEAVTARLVGVLFVSVTTGPWGAVATSAGGEESLKCEDLKVGQY 118

|||||||:|||||||

Db 1 MAAAWPSGPSAPDAVTARLVGVLFVSVTTGPWGAVATSAGGEESLKCEDLKVGQY 56

RESULT 10

ADA57043

ID ADA57043 standard; protein; 221 AA.

XX

AC ADA57043;

XX

DT 20-NOV-2003 (first entry)

XX

DE Human secreted protein #326.

XX

KW immunosuppressive; antiinflammatory; antiasthmatic; antiallergic;

KW cytostatic; cerebroprotective; neuroprotective; nootropic;

KW cardiovascular; antiarteriosclerotic; gene therapy;

KW human secreted protein; immune disorder; inflammation;

KW respiratory disorder; cancer; CNS disorder; neurodegenerative disorders;

KW inflammatory bowel disease; nephritis; Crohn's disease; asthma; allergy;

KW multiple sclerosis; ischaemic brain injury; Parkinson's disease;

KW Alzheimer's disease; atherosclerosis; myocarditis; chromosome mapping;

KW triple helix formation; antisense gene therapy; forensic biology.

XX

OS Homo sapiens.

XX

PN WO2002102994-A2.

XX

PD 27-DEC-2002.

XX

PF 19-MAR-2002; 2002WO-US008278.

XX

PR 21-MAR-2001; 2001US-0277340P.

PR 19-JUL-2001; 2001US-0306171P.

PR 13-NOV-2001; 2001US-0331287P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Rosen CA, Ruben SM;

XX

DR WPI; 2003-167512/16.

DR N-PSDB; ADA56147.

XX

PT New human secreted polypeptides and polynucleotides, useful for  
PT diagnosing, treating or preventing e.g. immune disorders, inflammatory  
PT conditions, respiratory disorders, cancers, CNS disorders, or  
PT neurodegenerative disorders.  
XX  
PS Claim 13; SEQ ID NO 1233; 1754pp; English.  
XX  
CC The invention relates to 592 new human secreted polypeptides useful for  
CC diagnosing, treating or preventing e.g. immune disorders, inflammatory  
CC conditions, respiratory disorders, cancers, CNS disorders, or  
CC neurodegenerative disorders, or polypeptides comprising an amino acid  
CC sequence at least 95% identical to the new sequences. The polypeptides,  
CC antibodies or antibody fragments that bind to the polypeptides, nucleic  
CC acids encoding the polypeptides, agonists or antagonists that binds to  
CC the polypeptide, are useful in preparing diagnostic or pharmaceutical  
CC compositions for diagnosing, treating or preventing an e.g. immune  
CC disorders, inflammatory conditions (e.g. inflammatory bowel disease,  
CC nephritis or Crohn's disease), respiratory disorders (e.g. asthma and  
CC allergy), cancers (e.g. gastric, ovarian or lung cancer), CNS disorders  
CC (e.g. multiple sclerosis or ischaemic brain injury), neurodegenerative  
CC disorders (e.g. Parkinson's disease or Alzheimer's disease), and  
CC cardiovascular disorders (e.g. atherosclerosis or myocarditis). The  
CC polynucleotides are useful for chromosome identification, chromosome  
CC mapping, for controlling gene expression through triple helix formation  
CC or antisense DNA or RNA, in gene therapy, for identifying individuals  
CC from minute biological samples, in forensic biology, and as hybridization  
CC probes. The polypeptides are useful for as molecular weight markers on  
CC sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)  
CC gels, to raise antibodies, for testing biological activities, and for  
CC treating or preventing neural disorders, immune system disorders,  
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,  
CC renal, proliferative and/or cancerous diseases. This sequence corresponds  
CC to one of the polypeptide of the invention. Note: The sequence data for  
CC this patent did form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 221 AA;

Query Match 14.0%; Score 201; DB 6; Length 221;  
Best Local Similarity 45.7%; Pred. No. 3.4e-12;  
Matches 53; Conservative 12; Mismatches 39; Indels 12; Gaps 5;

Qy 135 CTNYTA--HVSC----FPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAVA 188  
||| |: ||| :|| | | :| | | | | : | | : |:|  
Db 112 CTNSTSCMTVSCPRQRYPA-NCTVRD---HVVHCLGNRT-FPKMLYCNWTGGYKWSTALA 165  
Qy 189 LSLFLGWLGADRFYLGYPALGLLKFCITVGFCGIGSLIDFILISMQIVGPSDGSSYI 244  
||: || ||||| | | | : | || :||| :|| : |||:| ||  
Db 166 LSITLGGFGADRFYLGQWXEGLGKLFSFGGLGIWTLIDVLLIGVGYVGPADGSLYI 221

RESULT 11  
ABO14063  
ID ABO14063 standard; protein; 221 AA.  
XX  
AC ABO14063;

XX  
 DT 21-AUG-2003 (first entry)  
 XX  
 DE Novel human secreted protein #92.  
 XX  
 KW Human; secreted protein; cytostatic; neuroprotective; hepatotropic;  
 KW gene therapy; cancer; liver disorder; hepatitis; neural disorder;  
 KW Alzheimer's disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2003028003-A1.  
 XX  
 PD 06-FEB-2003.  
 XX  
 PF 12-OCT-2001; 2001US-00974879.  
 XX  
 PR 07-NOV-1997; 97US-0064900P.  
 PR 07-NOV-1997; 97US-0064908P.  
 PR 07-NOV-1997; 97US-0064911P.  
 PR 07-NOV-1997; 97US-0064912P.  
 PR 07-NOV-1997; 97US-0064983P.  
 PR 07-NOV-1997; 97US-0064984P.  
 PR 07-NOV-1997; 97US-0064985P.  
 PR 07-NOV-1997; 97US-0064987P.  
 PR 07-NOV-1997; 97US-0064988P.  
 PR 17-NOV-1997; 97US-0066089P.  
 PR 17-NOV-1997; 97US-0066090P.  
 PR 17-NOV-1997; 97US-0066094P.  
 PR 17-NOV-1997; 97US-0066095P.  
 PR 17-NOV-1997; 97US-0066100P.  
 PR 04-NOV-1998; 98WO-US023435.  
 PR 05-MAY-1999; 99US-00305736.  
 PR 13-OCT-2000; 2000US-0239893P.  
 PR 28-MAR-2001; 2001US-00818683.  
 XX  
 PA (ROSE/) ROSEN C A.  
 PA (FENG/) FENG P.  
 PA (RUBE/) RUBEN S M.  
 PA (EBNE/) EBNER R.  
 PA (OLSE/) OLSEN H S.  
 PA (NIJJ/) NI J.  
 PA (WEIY/) WEI Y.  
 PA (SOPP/) SOPPET D R.  
 PA (MOOR/) MOORE P A.  
 PA (KYAW/) KYAW H.  
 PA (LAFLE/) LAFLEUR D W.  
 PA (SHIY/) SHI Y.  
 PA (JANA/) JANAT F.  
 PA (ENDR/) ENDRESS G A.  
 PA (CART/) CARTER K C.  
 PA (BIRS/) BIRSE C E.  
 XX  
 PI Rosen CA, Feng P, Ruben SM, Ebner R, Olsen HS, Ni J, Wei Y;  
 PI Soppet DR, Moore PA, Kyaw H, Lafleur DW, Shi Y, Janat F;  
 PI Endress GA, Carter KC, Birse CE;  
 XX

DR WPI; 2003-479549/45.  
 DR N-PSDB; ACD18950.  
 XX  
 PT New nucleic acid molecule, useful for preparing a medicament for  
 PT preventing, treating or ameliorating a medical condition e.g., cancer,  
 PT liver disorders such as hepatitis or neural disorders such as Alzheimer's  
 PT disease.  
 XX  
 PS Claim 11; Page 387-388; 496pp; English.  
 XX  
 CC The invention describes a new isolated nucleic acid molecule comprising a  
 CC sequence having at least 95% identity with a sequence comprising: (a) a  
 CC polynucleotide (PN) fragment of a sequence comprising 420-3435 bp, or its  
 CC allelic variant; (b) a PN fragment of the cDNA sequence; (c) a PN  
 CC sequence encoding a polypeptide, or its fragment, domain, epitope or  
 CC species homologue; or (d) a PN that hybridises under stringent conditions  
 CC to any one of the sequences of (A)-(C). The nucleic acid is useful for  
 CC preparing a medicament for preventing, treating or ameliorating a medical  
 CC condition e.g., cancer, liver disorders such as hepatitis or neural  
 CC disorders such as Alzheimer's disease. This is the amino acid sequence of  
 CC a novel human secreted protein  
 XX  
 SQ Sequence 221 AA;

Query Match 14.0%; Score 201; DB 6; Length 221;  
 Best Local Similarity 45.7%; Pred. No. 3.4e-12;  
 Matches 53; Conservative 12; Mismatches 39; Indels 12; Gaps 5;

Qy 135 CTNYTA--HVSC----FPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAVA 188  
 ||| |: ||| :|| || :| | || | | : | || : |:|  
 Db 112 CTNSTSCMTVSCPRQRYPA-NCTVRD---HVHCLGNRT-FPKMLYCNWTGGYKWSTALA 165  
 Qy 189 LSLFLGWLGADRFYLGYPALGLLKFCTVGFCEGIGSLIDFILISMQIVGPSDGSSYI 244  
 ||: || ||||| | | | : | || :||| :|| : |||: || ||  
 Db 166 LSITLGGFGADRFYLGQWXEGLGKLFSFGGLGIWTLIDVLLIGVGYPADGSLYI 221

# RESULT 12

ABR47818

ID ABR47818 standard; protein; 221 AA.

XX

AC ABR47818;

XX

DT 12-JUN-2003 (first entry)

XX

DE Human secreted protein, SEQ ID 709.

XX

KW Cardiant; antiarrhythmic; antiarteriosclerotic; vasotropic; cytostatic;

KW vulnerary; antiinflammatory; nootropic; neuroprotective;

KW antiparkinsonian; gene therapy; human; cardiovascular disorder.

XX

OS Homo sapiens.

XX

PN WO200295010-A2.

XX

PD 28-NOV-2002.

XX

PF 19-MAR-2002; 2002WO-US009785.  
 XX  
 PR 21-MAR-2001; 2001US-0277340P.  
 PR 19-JUL-2001; 2001US-0306171P.  
 PR 13-NOV-2001; 2001US-0331287P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM;  
 XX  
 DR WPI; 2003-129429/12.  
 XX  
 PT Novel human secreted proteins, useful for detecting, preventing,  
 PT diagnosing, prognosticating, treating and/or ameliorating cardiovascular  
 PT disorders such as arrhythmia.  
 XX  
 PS Claim 13; SEQ ID NO 709; 1881pp; English.  
 XX  
 CC The present invention relates to novel human secreted proteins (ABR47633-  
 CC ABR48145) and their coding sequences (ACC50344-ACC50856). The proteins  
 CC and their coding sequences are useful for the preparation of a diagnostic  
 CC or pharmaceutical composition for diagnosing or treating a cardiovascular  
 CC disorder (e.g., arrhythmia, tachycardia, cardiac arrest, coronary  
 CC arteriosclerosis and myocardial ischaemia), neural disorders, immune  
 CC system disorders, muscular disorders, reproductive disorders,  
 CC gastrointestinal disorders, pulmonary disorders, renal disorders,  
 CC proliferative disorders and/or cancerous diseases and conditions, for  
 CC wound healing and epithelial cell proliferation, to treat inflammation or  
 CC infection, for treating thrombosis and arteriosclerosis, for treating or  
 CC preventing neural damage which occurs in neuronal disorders or  
 CC neurodegenerative conditions such as Alzheimer's disease and Parkinson's  
 CC disease, to enhance bone and periodontal regeneration and aid in tissue  
 CC transplants or bone grafts, to prevent skin aging or hair loss, to  
 CC stimulate growth and differentiation of haematopoietic cells and bone  
 CC marrow cells when used in combination with other cytokines, to maintain  
 CC organs before transplantation or for supporting cell culture of primary  
 CC tissues, to increase or decrease differentiation or proliferation of  
 CC embryonic stem cells, or to modulate mammalian characteristics or  
 CC metabolism. Note: The sequence data for this patent was published in  
 CC electronic format and is available from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 221 AA;

Query Match 14.0%; Score 201; DB 6; Length 221;  
 Best Local Similarity 45.7%; Pred. No. 3.4e-12;  
 Matches 53; Conservative 12; Mismatches 39; Indels 12; Gaps 5;

Qy 135 CTNYTA--HVSC----FPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAVA 188  
 ||| |: ||| :|| || :| | || || :| || : |:|  
 Db 112 CTNSTSCMTVSCPRQRYPA-NCTVRD---HVVHCLGNRT-FPKMLYCNWTGGYKWSTALA 165  
 Qy 189 LSLFLGWLGADRFYLGYPALGLLKFC TVGFCGIGSLIDFILISMQIVGPSDGSSYI 244  
 ||: || ||||| || | :| || :|| :|| : |||: || ||  
 Db 166 LSITLGGFGADRFYLGQWXEGLGKLFSFGGLGIWTLIDVLLIGVGVGPADGSLYI 221

RESULT 13

ABR00112

ID ABR00112 standard; protein; 221 AA.

XX

AC ABR00112;

XX

DT 03-APR-2003 (first entry)

XX

DE Human gene 102 encoded secreted protein HMEED18, SEQ ID NO:401.

XX

KW Human; secreted protein; digestive disorder; gastrointestinal disorder;  
 KW mouth; oesophagus; stomach; small intestine; large intestine; liver;  
 KW biliary tract; pancreas; cancer; tumour; hyperproliferative disorder;  
 KW immune disorder; inflammation; infection; wound healing; drug screening;  
 KW chromosome identification; chromosome mapping; cytostatic;  
 KW antiinflammatory; immunosuppressive; vulnerary; gene therapy.

XX

OS Homo sapiens.

XX

PN WO200276488-A1.

XX

PD 03-OCT-2002.

XX

PF 19-MAR-2002; 2002WO-US008276.

XX

PR 21-MAR-2001; 2001US-0277340P.

PR 19-JUL-2001; 2001US-0306171P.

PR 13-NOV-2001; 2001US-0331287P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Rosen CA, Ruben SM;

XX

DR WPI; 2003-029900/02.

DR N-PSDB; ABZ71291.

XX

PT New human secreted proteins and nucleic acids, useful for detecting,  
 PT preventing, diagnosing, prognosticating, treating and/or ameliorating  
 PT e.g. gastrointestinal diseases and disorders, or cancers.

XX

PS Claim 13; Page 1007; 1216pp; English.

XX

CC ABZ71190-ABZ71478 represent cDNAs corresponding to 178 human secreted  
 CC protein genes, and ABP00011-ABP00299 represent the proteins they encode.  
 CC ABZ71479-ABZ71540 represent human secreted protein genomic fragments. The  
 CC invention also encompasses antibodies specific for the secreted proteins,  
 CC the use of the secreted proteins in drug screening, and recombinant  
 CC vectors and host cells comprising a nucleic acid of the invention. The  
 CC secreted proteins, nucleic acids encoding them, antibodies or antibody  
 CC fragments specific for the secreted proteins, and modulators of protein  
 CC activity are useful for diagnosing, treating, ameliorating or preventing  
 CC digestive disorders. Such conditions include disorders of the mouth,  
 CC oesophagus, stomach, small intestine, large intestine, liver, biliary  
 CC tract and pancreas, and include cancers of these organs and tissues. The  
 CC secreted proteins and their nucleic acids may also be used in the  
 CC treatment of immune disorders, inflammation, infection,  
 CC hyperproliferative disorders, and to promote wound healing. Nucleic acids

CC of the invention may be used for chromosome identification, chromosome  
CC mapping, in gene therapy, for identifying individuals from minute  
CC biological samples, as hybridisation probes, and as molecular weight  
CC markers. The present sequence represents a human secreted protein of the  
CC invention

XX

SQ Sequence 221 AA;

Query Match 14.0%; Score 201; DB 6; Length 221;

Best Local Similarity 45.7%; Pred. No. 3.4e-12;

Matches 53; Conservative 12; Mismatches 39; Indels 12; Gaps 5;

```
Qy      135 CTNYTA--HVSC---FPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAVA 188
          ||| |: ||| :|| | | :|      | || | | : | || : |:|
Db      112 CTNSTSCMTVSCPRQRYPA-NCTVRD---HVVHCLGNRT-FPKMLYCNWTGGYKWSTALA 165

Qy      189 LSLFLGWLGA DRFYLGYPALGLLKFCTVGF CGIGSLIDFILISMQIVGPSDGSSYI 244
          ||: || ||||| || | : | || :|| :|| : |||: || ||
Db      166 LSITLGGFGADRFYLGQWXEGLGKLFSFGGLGIWTLIDVLLIGVGYVGPADGSLYI 221
```

RESULT 14

ADB91589

ID ADB91589 standard; protein; 221 AA.

XX

AC ADB91589;

XX

DT 04-DEC-2003 (first entry)

XX

DE Human secreted protein #SEQ ID 535.

XX

KW Secreted protein; gene therapy; antidiabetic; diabetes; human.

XX

OS Homo sapiens.

XX

PN WO2003004622-A2.

XX

PD 16-JAN-2003.

XX

PF 19-MAR-2002; 2002WO-US008124.

XX

PR 21-MAR-2001; 2001US-0277340P.

PR 19-JUL-2001; 2001US-0306171P.

PR 13-NOV-2001; 2001US-0331287P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Rosen CA, Ruben SM;

XX

DR WPI; 2003-229407/22.

XX

PT Nucleic acid encoding a human secreted protein is useful in diagnosing or  
PT treating diabetes or conditions related to diabetes.

XX

PS Claim 3; SEQ ID NO 535; 1537pp; English.

XX

CC The invention relates to isolated nucleic acid molecules ADB91065-



CC ADB91448 and ADB91835-ADB91911 encoding human secreted proteins ADB91449-  
 CC ADB91834. Also disclosed is a recombinant vector comprising a  
 CC polynucleotide of the invention, and a recombinant host cell comprising  
 CC the recombinant vector. The polypeptide of the invention is useful in  
 CC identifying a binding partner by contacting the polypeptide with a  
 CC binding partner, and determining whether the binding partner increases or  
 CC decreases activity of the polypeptide. The polypeptide, polynucleotide,  
 CC antibody or its fragment, agonist or antagonist are useful for preparing  
 CC a pharmaceutical composition for diagnosing or treating diabetes or  
 CC conditions related to diabetes. The present sequence is that of the human  
 CC immunoglobulin Fc portion used to generate fusion proteins, increasing  
 CC the stability of the fused protein as compared to the secreted protein  
 CC only. Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 221 AA;

Query Match 14.0%; Score 201; DB 7; Length 221;  
 Best Local Similarity 45.7%; Pred. No. 3.4e-12;  
 Matches 53; Conservative 12; Mismatches 39; Indels 12; Gaps 5;

Qy 135 CTNYTA--HVSC---FPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAVA 188  
 ||| |: ||| :|| || :| | || | | : | || : |:|  
 Db 112 CTNSTSCMTVSCPRQRYPA-NCTVRD---HVVHCLGNRT-FPKMLYCNWTGGYKWTALA 165  
 Qy 189 LSLFLGWLGADRFYLGYPALGLLKFCTVGFCGIGSLIDFILISMQIVGPSDGSSYI 244  
 ||: || ||||| || | : | || :|| :|| : ||: ||| ||  
 Db 166 LSITLGGFGADRFYLGQWXEGLGKLFSFGGLGIWTLIDVLLIGVGYVGPADGSLYI 221

# RESULT 15

ADC74204

ID ADC74204 standard; protein; 221 AA.

XX

AC ADC74204;

XX

DT 01-JAN-2004 (first entry)

XX

DE Human secreted protein - SEQ ID 837.

XX

KW antianaemic; antirheumatic; antiarthritic; antiinflammatory; antithyroid;  
 KW antidiabetic; immunosuppressive; dermatological; nephrotropic;  
 KW antiparkinsonian; neuroprotective; nootropic; antibacterial; virucide;  
 KW fungicide; antiparasitic; antiarteriosclerotic; vulnerary; cytostatic;  
 KW haemopoietic; haematologic; anaemia; autoimmune disorder;  
 KW rheumatoid arthritis; inflammation; Grave's disease; diabetes;  
 KW systemic lupus erythematosus; glomerulonephritis; neurodegenerative;  
 KW Parkinson's; Alzheimer's; wound; hyperproliferative; atherosclerosis;  
 KW cancer; bacterial; viral; fungal; parasitic infection; gene therapy;  
 KW human.

XX

OS Homo sapiens.

XX

PN WO2003038063-A2.

XX

PD 08-MAY-2003.

XX  
PF 19-MAR-2002; 2002WO-US008277.  
XX  
PR 21-MAR-2001; 2001US-0277340P.  
PR 19-JUL-2001; 2001US-0306171P.  
PR 13-NOV-2001; 2001US-0331287P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Ruben SM;  
XX  
DR WPI; 2003-430516/40.  
DR N-PSDB; ADC73589.  
XX  
PT New human secreted polypeptide for diagnosing, preventing or treating  
PT hematopoietic or hematologic disorders (e.g. anemia), autoimmune  
PT disorders (e.g. diabetes) or hyperproliferative disorders (e.g. cancer or  
PT atherosclerosis).  
XX  
PS Claim 16; SEQ ID NO 837; 2272pp; English.  
XX  
CC The invention relates to a novel human secreted polypeptide comprising a  
CC defined sequence given in the specification. The polypeptide, nucleic  
CC acid molecule, antibody, agonist or antagonist of the invention may be  
CC useful for preparing a composition for diagnosing or treating a  
CC haemopoietic or haematologic disorder such as anaemia, autoimmune  
CC disorders such as rheumatoid arthritis, inflammation, Grave's disease,  
CC diabetes, systemic lupus erythematosus or glomerulonephritis,  
CC neurodegenerative disorders including Parkinson's disease and Alzheimer's  
CC disease, wounds and hyperproliferative disorders including  
CC atherosclerosis or cancer, as well as bacterial, viral, fungal or  
CC parasitic infections. The polypeptide may also be used during gene  
CC therapy procedures and for identifying a binding partner by contacting  
CC the polypeptide with a binding partner and determining whether the  
CC binding partner increases or decreases the activity of the polypeptide.  
CC The current sequence is that of the human secreted protein of the  
CC invention.  
XX  
SQ Sequence 221 AA;

Query Match 14.0%; Score 201; DB 7; Length 221;  
Best Local Similarity 45.7%; Pred. No. 3.4e-12;  
Matches 53; Conservative 12; Mismatches 39; Indels 12; Gaps 5;

QY 135 CTNYTA--HVSC----FPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAVA 188  
||| |: ||| :|| || :| | || || :| || : |:|  
Db 112 CTNSTSCMTVSCPRQRYPA-NCTVRD----HVVHCLGNRT-FPKMLYCJWTGGYKWSTALA 165  
QY 189 LSLFLGWLGADRFYLGYPALGLLKFCVTGFCGIGSLIDFILISMQIVGPSDGSSYI 244  
||: || ||||| || | :| || :||| :|| : |||:| ||  
Db 166 LSITLGGFGADRFYLGQWXEGLGKLFSFGGLGIWTLIDVLLIGVGYPADGSLYI 221

Search completed: March 4, 2004, 10:24:13  
Job time : 99 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2004, 10:22:25 ; Search time 43 Seconds  
(without alignments)  
322.962 Million cell updates/sec

Title: US-09-852-100B-2  
Perfect score: 1439  
Sequence: 1 MHILKGSPNVIPRAHGQKNT.....TRLTRLSITNETFRKTQLYP 269

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	85.5	5.9	1023	4	US-10-164-595-20	Sequence 20, Appl
2	85	5.9	310	2	US-08-414-657D-45	Sequence 45, Appl
3	85	5.9	338	2	US-08-414-657D-42	Sequence 42, Appl
4	85	5.9	338	2	US-08-414-657D-43	Sequence 43, Appl
5	85	5.9	338	4	US-09-135-080-4	Sequence 4, Appli
6	84	5.8	258	4	US-09-328-352-4253	Sequence 4253, Ap
7	83	5.8	764	2	US-08-177-109A-2	Sequence 2, Appli
8	83	5.8	764	2	US-08-687-706-2	Sequence 2, Appli
9	82.5	5.7	338	4	US-09-976-594-404	Sequence 404, App
10	81	5.6	797	3	US-09-182-728A-2	Sequence 2, Appli
11	81	5.6	797	4	US-09-795-232-2	Sequence 2, Appli
12	80.5	5.6	150	4	US-09-252-991A-16958	Sequence 16958, A

13	80.5	5.6	304	2	US-08-414-657D-44	Sequence 44, Appl
14	80.5	5.6	325	2	US-08-414-657D-2	Sequence 2, Appli
15	80.5	5.6	325	2	US-08-414-657D-41	Sequence 41, Appl
16	80.5	5.6	325	4	US-09-135-080-2	Sequence 2, Appli
17	80.5	5.6	456	3	US-09-058-389A-4	Sequence 4, Appli
18	80.5	5.6	456	4	US-09-611-781-4	Sequence 4, Appli
19	80.5	5.6	1070	4	US-10-164-595-18	Sequence 18, Appl
20	80.5	5.6	1073	4	US-10-164-595-22	Sequence 22, Appl
21	80.5	5.6	1237	1	US-08-241-853-2	Sequence 2, Appli
22	80.5	5.6	1237	2	US-08-850-917-2	Sequence 2, Appli
23	80	5.6	328	4	US-09-253-316-28	Sequence 28, Appl
24	80	5.6	660	3	US-09-181-706-8	Sequence 8, Appli
25	80	5.6	660	3	US-09-458-791-8	Sequence 8, Appli
26	80	5.6	660	3	US-09-459-066-8	Sequence 8, Appli
27	80	5.6	660	4	US-09-459-065-8	Sequence 8, Appli
28	79.5	5.5	354	2	US-08-700-013B-9	Sequence 9, Appli
29	79.5	5.5	797	2	US-08-700-013B-21	Sequence 21, Appl
30	79.5	5.5	927	4	US-09-328-352-7922	Sequence 7922, Ap
31	79	5.5	797	2	US-08-700-013B-19	Sequence 19, Appl
32	79	5.5	797	4	US-09-191-468-124	Sequence 124, App
33	78.5	5.5	232	4	US-09-252-991A-19777	Sequence 19777, A
34	78.5	5.5	338	2	US-08-414-657D-60	Sequence 60, Appl
35	78.5	5.5	338	4	US-09-135-080-8	Sequence 8, Appli
36	78	5.4	493	4	US-09-112-498A-2	Sequence 2, Appli
37	77	5.4	74	4	US-09-543-681A-6319	Sequence 6319, Ap
38	77	5.4	357	4	US-09-252-991A-31718	Sequence 31718, A
39	77	5.4	379	4	US-09-252-991A-26357	Sequence 26357, A
40	77	5.4	490	4	US-09-489-039A-9610	Sequence 9610, Ap
41	77	5.4	797	4	US-09-191-468-120	Sequence 120, App
42	77	5.4	797	4	US-09-191-468-122	Sequence 122, App
43	76	5.3	467	4	US-09-411-132A-4	Sequence 4, Appli
44	76	5.3	488	4	US-09-252-991A-26323	Sequence 26323, A
45	76	5.3	525	1	US-08-356-340-2	Sequence 2, Appli

#### ALIGNMENTS

##### RESULT 1

US-10-164-595-20

; Sequence 20, Application US/10164595

; Patent No. 6657054

; GENERAL INFORMATION:

; APPLICANT: OriGene Technologies, Inc

; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides

; FILE REFERENCE: 1U 103 R1

; CURRENT APPLICATION NUMBER: US/10/164,595

; CURRENT FILING DATE: 2002-06-10

; NUMBER OF SEQ ID NOS: 80

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 20

; LENGTH: 1023

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-164-595-20

Query Match

5.9%; Score 85.5; DB 4; Length 1023;

Best Local Similarity 20.0%; Pred. No. 10;  
Matches 68; Conservative 32; Mismatches 103; Indels 137; Gaps 14;

```

Qy      3 ILKGSPNVIP-----RAHGQ-----KNTRRDGTGLYPMRGPFKNLALLPFSLP LL- 47
      : | |:| | | | | : | | :| | | | : | : | |
Db      328 VSSGKPSVAPKPAANRASGEWDSGTENRLKVTSKEGLTPYP---PLQEAGSIPVTKPELP 384

Qy      48 -----GGG----GSGSGEKV-----SVSKMAAAWPSGPSAPE 75
      | | | | | | | | | | | | | | | | | | | |
Db      385 KKPNPGLIRSVNPEIPGRGPLAESSDSGKKVPTPAPRPLLLKKSVSSENPTYPSAPLKPV 444

Qy      76 AVTARLVG-----VLWFSVTTGPGWGA VATSA 102
      | | | | | | | | | | | | | | | | | | | |
Db      445 TVPPRLAGASQAKAYKSLGEGPPANPPVPVLQSKPLVDIDLISFDDDLPTPSGNLAEEES 504

Qy      103 GGEESL-----KCEDLK---VGQYICKDPKINDATQEPVNCTNYTAHVSCFPAPNIT 151
      | | : | | :| | | : | : | | | | | | | |
Db      505 VGSEMLDPFQLPAKTEPIKERA VQPAPTRKPTVIRIPAKPGKC----LHEDPQSPPLP 560

Qy      152 CKDSSGN-----ETHFTGNEVGFFK---PISCRNVNGYSYKVAVALS 190
      : | | | | | | | | | | | | | | | | | |
Db      561 AEKPIGNTFSTVSGKLSNVERTRNLESNHPGQTGGFVRVPPRLPPRPVNGHLIMTTI--- 617

Qy      191 LFLGWL GADRFYLGYPALGLLKFCTVGFCGIGSLIDFILI 230
      ||: | | | | | | | | | | | | | | | |
Db      618 LFMSCSARAR-----MGFTGIVHILRFKLL 642

```

# RESULT 2

US-08-414-657D-45

; Sequence 45, Application US/08414657D

; Patent No. 5861283

## ; GENERAL INFORMATION:

; APPLICANT: Levitt, Pat

; APPLICANT: Pimenta, Aurea

; APPLICANT: Fischer, Itzhak

; APPLICANT: Zhukareva, Victoria

; TITLE OF INVENTION: Limbic System-Associated Membrane

; TITLE OF INVENTION: Protein and DNA

; NUMBER OF SEQUENCES: 60

## ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Dechert Price & Rhoads

; STREET: 997 Lenox Drive, Building 3, Suite 210

; CITY: Lawrenceville

; STATE: NJ

; COUNTRY: USA

; ZIP: 08543

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ for Windows Version 2.0

## ; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/414,657D

; FILING DATE: 31-MAR-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 317743-102
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 310 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-414-657D-45

```

```

Query Match          5.9%; Score 85; DB 2; Length 310;
Best Local Similarity 27.7%; Pred. No. 2.1;
Matches 36; Conservative 15; Mismatches 47; Indels 32; Gaps 7;

```

```

Qy      101 SAGGEESLKCEDLKVG----QYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSS 156
      : | : |||| |      :: | :|| |      : |      | :| ::
Db      202 TTGRQASLKCEASAVPAPDFEWYRDDTRINSANGLEIKSTE---GQSSLTVTNVT-EEHY 257

Qy      157 GNETHFTGNEVG-----FFKPISCRNVNGYSYKVAVALSLFLGWLGADR FYLGYPALG 209
      || | | : :|      | :| | | :|| | :|| |      ||      |
Db      258 GNYTCVAANKLGVTNASLVLFVRPGSVRGING-SISLAVPL-----WL-----LA 300

Qy      210 LLKFCTVGFC 219
      || : |
Db      301 ASLFCLLSKC 310

```

# RESULT 3

US-08-414-657D-42

; Sequence 42, Application US/08414657D

; Patent No. 5861283

; GENERAL INFORMATION:

; APPLICANT: Levitt, Pat

; APPLICANT: Pimenta, Aurea

; APPLICANT: Fischer, Itzhak

; APPLICANT: Zhukareva, Victoria

; TITLE OF INVENTION: Limbic System-Associated Membrane

; TITLE OF INVENTION: Protein and DNA

; NUMBER OF SEQUENCES: 60

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Dechert Price & Rhoads

; STREET: 997 Lenox Drive, Building 3, Suite 210

; CITY: Lawrenceville

; STATE: NJ

; COUNTRY: USA

; ZIP: 08543

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

```

;   OPERATING SYSTEM:  DOS
;   SOFTWARE:  FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  US/08/414,657D
;   FILING DATE:  31-MAR-1995
;   CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME:  Bloom, Allen
;   REGISTRATION NUMBER:  29,135
;   REFERENCE/DOCKET NUMBER:  317743-102
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  609-520-3214
;   TELEFAX:  609-520-3259
;   TELEX:
;   INFORMATION FOR SEQ ID NO:  42:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH:  338 amino acids
;   TYPE:  amino acid
;   STRANDEDNESS:  single
;   TOPOLOGY:  linear
US-08-414-657D-42

```

```

Query Match          5.9%;  Score 85;  DB 2;  Length 338;
Best Local Similarity  27.7%;  Pred. No. 2.4;
Matches  36;  Conservative  15;  Mismatches  47;  Indels  32;  Gaps  7;

```

```

Qy      101 SAGGEESLKCEDLKVG----QYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSS 156
      : | : |||| |      :: | :|| | : | |      | :| ::
Db      230 TTGRQASLKCEASAVPAPDFEWYRDDTRINSANGLEIKSTE---GQSSLTVTNVT-EEHY 285

Qy      157 GNETHFTGNEVG-----FFKPISCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALG 209
      || |      |::|      |:|| | :|| | :|| |      ||      |
Db      286 GNYTCVAANKLGVTNASLVLFRPGSVRGING-SISLAVPL-----WL-----LA 328

Qy      210 LLKFCTVGFC 219
      || : |
Db      329 ASLFCLLSKC 338

```

#### RESULT 4

US-08-414-657D-43

```

; Sequence 43, Application US/08414657D
; Patent No. 5861283
; GENERAL INFORMATION:
;   APPLICANT:  Levitt, Pat
;   APPLICANT:  Pimenta, Aurea
;   APPLICANT:  Fischer, Itzhak
;   APPLICANT:  Zhukareva, Victoria
;   TITLE OF INVENTION:  Limbic System-Associated Membrane
;   TITLE OF INVENTION:  Protein and DNA
;   NUMBER OF SEQUENCES:  60
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE:  Dechert Price & Rhoads
;   STREET:  997 Lenox Drive, Building 3, Suite 210

```

```

; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/414,657D
; FILING DATE: 31-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 317743-102
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 338 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-414-657D-43

```

```

Query Match          5.9%; Score 85; DB 2; Length 338;
Best Local Similarity 27.7%; Pred. No. 2.4;
Matches 36; Conservative 15; Mismatches 47; Indels 32; Gaps 7;

```

```

Qy      101 SAGGEESLKCEDLKVG---QYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSS 156
      : | : |||| | : : | : || | : | | :| : :
Db      230 TTGRQASLKCEASAVPAPDFEYWRDDTRINSANGLEIKSTE---GQSSLTVTNVT-EEHY 285

Qy      157 GNETHFTGNEVG-----FFKPISCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALG 209
      || | | : : | | : | | | : || | : || | |
Db      286 GNYTCVAANKLGVTNASLVLFRPGSVRGING-SISLAVPL-----WL-----LA 328

Qy      210 LLKFCTVGFC 219
      || : |
Db      329 ASLFCLLSKC 338

```

# RESULT 5

US-09-135-080-4

; Sequence 4, Application US/09135080

; Patent No. 6423827

; GENERAL INFORMATION:

; APPLICANT: Levitt, Pat R.

; APPLICANT: Pimenta, Aurea

; APPLICANT: Fischer, Itzhak



```

; APPLICANT: Zhukareva, Victoria
; TITLE OF INVENTION: Limbic System-Associated Membrane
; TITLE OF INVENTION: Protein and DNA
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135,080
; FILING DATE: 17-AUG-1998
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/414,657
; FILING DATE: 31-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 317743-102A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-620-3214
; TELEFAX: 609-620-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 338 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-135-080-4

```

```

Query Match          5.9%; Score 85; DB 4; Length 338;
Best Local Similarity 27.7%; Pred. No. 2.4;
Matches 36; Conservative 15; Mismatches 47; Indels 32; Gaps 7;

```

```

QY      101 SAGGEESLKCEDLKVG----QYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSS 156
      : | : |||| |      :: | :|| | : |      | :| ::
Db      230 TTGRQASLKCEASAVPAPDFEWYRDDTRINSANGLEIKSTE---GQSSLTVTNVT-EEHY 285

QY      157 GNETHFTGNEVG-----FFKPISCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALG 209
      || | |::|      |:| | | :|| | :|| |      ||      |
Db      286 GNYTCVAANKLGVTNASLVLFVRPGSVRGING-SISLAVPL-----WL-----LA 328

QY      210 LLKFCTVGFC 219
      || : |
Db      329 ASLFCLLSKC 338

```

RESULT 6

US-09-328-352-4253  
 ; Sequence 4253, Application US/09328352  
 ; Patent No. 6562958  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary L. Breton et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 ACINETOBACTER  
 ; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC99-03PA  
 ; CURRENT APPLICATION NUMBER: US/09/328,352  
 ; CURRENT FILING DATE: 1999-06-04  
 ; NUMBER OF SEQ ID NOS: 8252  
 ; SEQ ID NO 4253  
 ; LENGTH: 258  
 ; TYPE: PRT  
 ; ORGANISM: Acinetobacter baumannii  
 US-09-328-352-4253

Query Match 5.8%; Score 84; DB 4; Length 258;  
 Best Local Similarity 23.5%; Pred. No. 2;  
 Matches 52; Conservative 24; Mismatches 93; Indels 52; Gaps 10;

```

Qy      78 TARLVGVL---WFVSVTTGPW-----GAVATSAGGEESLKCEDLKV-----G 116
      | | | | | : ||: | | |:: | |
Db      57 TTGLYGPLNVEWTTTRLERGPYWSEKIDEKGTFFRGAPGSISIRSPDYPSPGQPAATDGG 116

Qy     117 QYICKDPKINDATQEPVNCTNY--TAHVSC-FPAPNITCKDSSGNETHFTGNEVGFFKPI 173
      |: |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     117 FYLPKDPK-----EPVKIYRYFTTKAVPVEVPSDNVTC-----NTLAYTKEP 158

Qy     174 SCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCTVGFCEGIGSLIDFILISMQ 233
      : | : | : | : | : | | : | | | | | | | | | | | | | | | | |
Db     159 ASHKVSLVSFATAGTVGGVTGAIIGKNFSSGNMSYQATGAGAAGGAIGGLIVAAIINAE 218

Qy     234 ---IVG--PSDGSSYIIDYYGTRLRLSITNETFRKTQLYP 269
      |: | | | |::: : | | | : : | |
Db     219 VGKIIIGGLPIKESSFM-----EKLRELGAkrePLKQISLLP 254

```

# RESULT 7

US-08-177-109A-2  
 ; Sequence 2, Application US/08177109A  
 ; Patent No. 5869615  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dennis E. Hourcade and Teresa J. Oglesby  
 ; TITLE OF INVENTION: MODIFIED COMPLEMENT PROTEASES  
 ; NUMBER OF SEQUENCES: 62  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Patrea L. Pabst  
 ; STREET: 2800 One Atlantic Center  
 ; STREET: 1201 West Peachtree Street  
 ; CITY: Atlanta  
 ; STATE: Georgia  
 ; COUNTRY: USA  
 ; ZIP: 30309-3450  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk

```

;   COMPUTER:  IBM PC compatible
;   OPERATING SYSTEM:  PC-DOS/MS-DOS
;   SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  US/08/177,109A
;   FILING DATE:  03-JAN-1994
;   CLASSIFICATION:  514
;   ATTORNEY/AGENT INFORMATION:
;   NAME:  Pabst, Patrea L.
;   REGISTRATION NUMBER:  31,284
;   REFERENCE/DOCKET NUMBER:  WU 107
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  (404) 873-8794
;   TELEFAX:  (404) 873-8795
;   INFORMATION FOR SEQ ID NO:  2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH:  764 amino acids
;   TYPE:  amino acid
;   TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
;   HYPOTHETICAL:  NO
US-08-177-109A-2

```

```

Query Match          5.8%;  Score 83;  DB 2;  Length 764;
Best Local Similarity 24.1%;  Pred. No. 12;
Matches  49;  Conservative  21;  Mismatches  71;  Indels  62;  Gaps  12;

```

```

Qy      24 GTGLYPMRGPFKNLALLPFSLP LLGGG-----GSGSGEKVSV----- 60
      |: || |      | |: || | || ||      || | | | :
Db      2 GSNLSP-----QLCLMPFILGLLSGGVTTTPWSLAQPQGSCLSGVEIKGGSFRLLQEG 55

Qy      61 SKMAAAWPSG--PSAPEAVTARLVGV LW FVS VTTGPGW GAVATS-----AGGEESLKC-- 110
      :   ||| | :   || |      : || | : |      :: |
Db      56 QALEYVCPSG FYPYPVQTRTCR-----STGSWSTLKTQDQKTVRKAECRAIHCPR 105

Qy      111 -EDLKVGQYICKDEPKINDATQEPVNC-TNYTAHVSCFPAPNITCKDSS--GNETHFTGNE 166
      | : |: | : | | : : : | || | | | | | : : : | |
Db      106 PHDFENGEYWPRSPYYNVSD EISFHCYDGYTLRGSA---NRTCQVNGRWSGQTAICDNG 161

Qy      167 VGFFK----PISCRNVNGYSYKV 185
      |:      || | | | | : :
Db      162 AGYCSNPGIPIGTRKV-GSQYRL 183

```

# RESULT 8

US-08-687-706-2

```

; Sequence 2, Application US/08687706
; Patent No. 5928892
; GENERAL INFORMATION:
;   APPLICANT:  Dennis E. Hourcade and Teresa J. Oglesby
;   TITLE OF INVENTION:  MODIFIED COMPLEMENT PROTEASES
;   NUMBER OF SEQUENCES:  62
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE:  Patrea L. Pabst
;   STREET:  2800 One Atlantic Center
;   STREET:  1201 West Peachtree Street
;   CITY:  Atlanta

```

```

; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,706
; FILING DATE: 26-JUL-1996
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/177,109
; FILING DATE: 03-JAN-1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: WU 107 DIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 764 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-08-687-706-2

```

```

Query Match          5.8%; Score 83; DB 2; Length 764;
Best Local Similarity 24.1%; Pred. No. 12;
Matches 49; Conservative 21; Mismatches 71; Indels 62; Gaps 12;

```

```

Qy      24 GTGLYPMRGPFKNLALLPFSLPLLGGG-----GSGSGEKVSV----- 60
      | : | |      | | : | | | | |      | | | | | :
Db      2 GSNLSP-----QLCLMPFILGLLSGGVTTTPWSLAQPQGSCLLEGVEIKGGSFRLQEG 55

Qy      61 SKMAAAWPSG--PSAPEAVTARLVGVLFVSVTTGPWGAVATS-----AGGEESLKC-- 110
      :   | | | | :   : | |      : | | : |      : : |
Db      56 QALEYVCPSGFYYPVQTRTCR-----STGSWSTLKTQDQKTVRKAECRAIHCPR 105

Qy      111 -EDLKVGQYICKDPKINDATQEPVNC-TNYTAHVSCFPAPNITCKDSS--GNETHFTGNE 166
      | : | : | : | | : : : | | | | | | | | : : | |
Db      106 PHDFENGGEYWPRSPYYNVSDAISFHCDYGTTLRGSA----NRTCQVNGRWSGQTAICDNG 161

Qy      167 VGFFK----PISCRNVNGYSYKV 185
      | :      | | | | | | : :
Db      162 AGYCSNPGIPIGTRKV-GSQYRL 183

```

# RESULT 9

US-09-976-594-404

; Sequence 404, Application US/09976594

; Patent No. 6673549



; SEQ ID NO 2  
; LENGTH: 797  
; TYPE: PRT  
; ORGANISM: HOMO SAPIENS  
US-09-182-728A-2

Query Match 5.6%; Score 81; DB 3; Length 797;  
Best Local Similarity 23.9%; Pred. No. 21;  
Matches 47; Conservative 32; Mismatches 80; Indels 38; Gaps 11;

```
Qy      87 FVSVTTGPWGAVATSAGGEESLKCED---LKVGQYICKD-PKINDATQEPVNCTNYTAHV 142
      |||| |||: | |:| | : | ||| | | :
Db      302 FVSVL--PWGSCNNPWNTPE---CKDKTKLLLDSCVISDHPKI-----QIKNSTFCM 348

Qy      143 SCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSY----KVAVALSLFLGWLGA 198
      : :| : | |:| |:| :|| : | | : :|| ||| |:
Db      349 TAYPNVTMVNFTSQANKTFVSGSE-EYFKYFVLKISAGIEYPGEIRWPLALCLFLAWV-- 405

Qy      199 DRFYLGYPAL-----GLLKFTVGF CGIGSLIDFILISMQIVGPSDGSSYIIDYYGTR 251
      : | :| | : : | | : |: : : : | | | :
Db      406 ----IVYASLAKGIKTSKGVVYFTATFPYV-VLVILLIRGVTLPGAGAGIWIYFITPKWEK 460

Qy      252 LTRLSITNETFRKTQLY 268
      || :: : ||::
Db      461 LTNATVWKDA--ATQIF 475
```

RESULT 11

US-09-795-232-2

; Sequence 2, Application US/09795232  
; Patent No. 6426405  
; GENERAL INFORMATION:  
; APPLICANT: Anthony M. Brown  
; APPLICANT: Conrad Gerald Chapman  
; APPLICANT: Israel Simon Gloger  
; APPLICANT: Joanne Rachel Evans  
; APPLICANT: William Cairns  
; APPLICANT: Hugh Jonathan Herdon  
; TITLE OF INVENTION: NOVEL COMPOUNDS  
; FILE REFERENCE: GP-30176-D1  
; CURRENT APPLICATION NUMBER: US/09/795,232  
; CURRENT FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: 09/182,728  
; PRIOR FILING DATE: 1998-10-29  
; PRIOR APPLICATION NUMBER: 9818890.7  
; PRIOR FILING DATE: 1998-08-28  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 797  
; TYPE: PRT  
; ORGANISM: HOMO SAPIENS  
US-09-795-232-2

Query Match 5.6%; Score 81; DB 4; Length 797;  
Best Local Similarity 23.9%; Pred. No. 21;  
Matches 47; Conservative 32; Mismatches 80; Indels 38; Gaps 11;

Qy 87 FVSVTTGPWGAVATSAGGEESLKCED---LKVGYICKD-PKINDATQEPVNCTNYTAHV 142  
 ||| | | : | : | : | | | | :  
 Db 302 FVSVL--PWGSCNNPWNTPE---CKDKTKLLDSCVISDHPKI-----QIKNSTFCM 348

Qy 143 SCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSY----KVAVALSLFLGWLGA 198  
 : : | : | : | : | : | : | : | : | : | :  
 Db 349 TAYPNVTMVNFTSQANKTFVSGSE-EYFKYFVLKISAGIEYPGEIRWPLALCLFLAWV-- 405

Qy 199 DRFYLGYPAL-----GLLKFTVGFCGIGSLIDFILISMQIVGPSDGSSYIIDYYGTR 251  
 : | : | | : : | | : | : : : : | | | :  
 Db 406 ----IVYASLAKGIKTSQKVVYFTATFPYV-VLVILLIRGVTLPGAGAGIWYFITPKWEK 460

Qy 252 LTRLSITNETFRKTQLY 268  
 || : : : || : :  
 Db 461 LTNATVWKDA--ATQIF 475

# RESULT 12

US-09-252-991A-16958  
 ; Sequence 16958, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 16958  
 ; LENGTH: 150  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-16958

Query Match 5.6%; Score 80.5; DB 4; Length 150;  
 Best Local Similarity 40.0%; Pred. No. 2.1;  
 Matches 20; Conservative 7; Mismatches 22; Indels 1; Gaps 1;

Qy 181 YSYKVAVALSLFLGWLGADRFYLGYPALGLLKFTVGFCGIGSLIDFILI 230  
 : | : | : | : | | | | : | | | | : | | |  
 Db 24 HSKAIGYLLWIF-GFTGSHRFYYGKPITGTIWFFTFGLFFIGWIIDLFLI 72

# RESULT 13

US-08-414-657D-44  
 ; Sequence 44, Application US/08414657D  
 ; Patent No. 5861283  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Levitt, Pat  
 ; APPLICANT: Pimenta, Aurea

```

; APPLICANT: Fischer, Itzhak
; APPLICANT: Zhukareva, Victoria
; TITLE OF INVENTION: Limbic System-Associated Membrane
; TITLE OF INVENTION: Protein and DNA
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/414,657D
; FILING DATE: 31-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 317743-102
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 304 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-414-657D-44

```

```

Query Match          5.6%; Score 80.5; DB 2; Length 304;
Best Local Similarity 29.9%; Pred. No. 5.8;
Matches 32; Conservative 14; Mismatches 40; Indels 21; Gaps 6;

```

```

Qy      101 SAGGEESLKCEDLKVG----QYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSS 156
      : | : |||| | : : | :|| | : | | :| : :
Db      202 TTGRQASLKCEASAVPAPDFEWYRDDTRINSANGLEIKSTE---GQSSLTVTNVT-EEHY 257

Qy      157 GNETHFTGNEVG-----FFKPISCRNVNGYSYKVAVALSLFLGWL 196
      || | | : : | :| | | :|| | :|| | ||
Db      258 GNYTCVAANKLGVTNASLVLFRPGSVRGING-SISLAVPL-----WL 298

```

```

RESULT 14
US-08-414-657D-2
; Sequence 2, Application US/08414657D
; Patent No. 5861283

```



```

; GENERAL INFORMATION:
;   APPLICANT:  Levitt, Pat
;   APPLICANT:  Pimenta, Aurea
;   APPLICANT:  Fischer, Itzhak
;   APPLICANT:  Zhukareva, Victoria
;   TITLE OF INVENTION:  Limbic System-Associated Membrane
;   TITLE OF INVENTION:  Protein and DNA
;   NUMBER OF SEQUENCES:  60
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Dechert Price & Rhoads
;     STREET:    997 Lenox Drive, Building 3, Suite 210
;     CITY:      Lawrenceville
;     STATE:     NJ
;     COUNTRY:   USA
;     ZIP:       08543
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Diskette
;     COMPUTER:     IBM Compatible
;     OPERATING SYSTEM:  DOS
;     SOFTWARE:     FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/414,657D
;     FILING DATE:       31-MAR-1995
;     CLASSIFICATION:    435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:
;     FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Bloom, Allen
;     REGISTRATION NUMBER:  29,135
;     REFERENCE/DOCKET NUMBER:  317743-102
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  609-520-3214
;     TELEFAX:   609-520-3259
;     TELEX:
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  325 amino acids
;       TYPE:    amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
;     FRAGMENT TYPE:  internal
US-08-414-657D-2

```

```

Query Match          5.6%;  Score 80.5;  DB 2;  Length 325;
Best Local Similarity 29.9%;  Pred. No. 6.4;
Matches 32;  Conservative 14;  Mismatches 40;  Indels 21;  Gaps 6;
Qy      101 SAGGEESLKCEDLKVG----QYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSS 156
      : | : ||||| |      :: | :|| |      : |      |      |:| ::
Db      223 TTGRQASLKCEASAVPAPDFEWYRDDTRINSANGLEIKSTE---GQSSLTVTNVT-EEHY 278
Qy      157 GNETHFTGNEVG-----FFKPISCRNVNGYSYKVAVALSLFLGWL 196
      || |      |:| |      |:| | | :|| |      :|| |      ||
Db      279 GNYTCVAANKLGVTNASLVLFRPGSVRGING-SISLAVPL-----WL 319

```

RESULT 15

US-08-414-657D-41  
; Sequence 41, Application US/08414657D  
; Patent No. 5861283  
; GENERAL INFORMATION:  
; APPLICANT: Levitt, Pat  
; APPLICANT: Pimenta, Aurea  
; APPLICANT: Fischer, Itzhak  
; APPLICANT: Zhukareva, Victoria  
; TITLE OF INVENTION: Limbic System-Associated Membrane  
; TITLE OF INVENTION: Protein and DNA  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dechert Price & Rhoads  
; STREET: 997 Lenox Drive, Building 3, Suite 210  
; CITY: Lawrenceville  
; STATE: NJ  
; COUNTRY: USA  
; ZIP: 08543  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/414,657D  
; FILING DATE: 31-MAR-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bloom, Allen  
; REGISTRATION NUMBER: 29,135  
; REFERENCE/DOCKET NUMBER: 317743-102  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 609-520-3214  
; TELEFAX: 609-520-3259  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 41:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 325 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

US-08-414-657D-41

Query Match 5.6%; Score 80.5; DB 2; Length 325;  
Best Local Similarity 29.9%; Pred. No. 6.4;  
Matches 32; Conservative 14; Mismatches 40; Indels 21; Gaps 6  
Qy 101 SAGGEESLKCEDLKVG----QYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSS 156  
: | : |||| | : : | :|| | : | | :| :  
Db 223 TTGRQASLKCEASAVPAPDFEYRDDTRINSANGLEIKSTE---GQSSLTVTNVT-EEHY 278  
QY 157 GNETHFTGNEVG-----FFKPISCRNVNGYSYKVAVALSLFLGWL 196  
|| | |::| | :| | | :|| | :|| | ||  
Db 279 GNYTCVAANKLGVTNASLVLFRPGSVRGING-SISLAVPL-----WL 319

Search completed: March 4, 2004, 10:28:43  
Job time : 44 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2004, 10:16:25 ; Search time 44 Seconds  
(without alignments)  
588.080 Million cell updates/sec

Title: US-09-852-100B-2  
Perfect score: 1439  
Sequence: 1 MHILKGSPNVIPRAHGQKNT.....TRLTRLSITNETFRKTQLYP 269

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query	Match Length			
1	167.5	11.6	573	2	S44605	C02F5.3 protein -
2	159.5	11.1	753	2	T28787	hypothetical prote
3	95	6.6	309	2	H75286	hypothetical prote
4	92	6.4	456	2	S55661	hypothetical prote
5	91	6.3	1324	1	VGIH59	E2 glycoprotein pr
6	89.5	6.2	690	2	T08604	hypothetical prote
7	87.5	6.1	547	2	H75632	Na(+)-linked D-ala
8	87.5	6.1	1207	2	T23754	hypothetical prote
9	85.5	5.9	348	1	S20911	alcohol dehydrogen
10	84	5.8	446	2	T35005	probable integral
11	83.5	5.8	487	2	C70574	probable aroP2 pro
12	83	5.8	137	2	B75447	hypothetical prote
13	83	5.8	348	1	S32521	alcohol dehydrogen

14	83	5.8	764	1	BBHU	complement factor
15	83	5.8	1274	2	T10729	transferrin-like p
16	82.5	5.7	338	2	JC4776	limbic-system-asso
17	82	5.7	409	2	T03788	laccase (EC 1.10.3
18	81.5	5.7	504	2	T37164	probable monooxyge
19	81.5	5.7	1266	2	AG2695	conserved hypothet
20	81.5	5.7	1266	2	G97477	hypothetical prote
21	81	5.6	443	2	G90388	thermopsine precur
22	81	5.6	513	2	A81850	probable integral
23	81	5.6	513	2	C81092	oxalate/formate an
24	81	5.6	758	2	D87369	beta-D-glucosidase
25	81	5.6	1176	2	T49482	hypothetical prote
26	80.5	5.6	134	2	D83595	hypothetical prote
27	80.5	5.6	223	2	F72803	gp32 protein - Myc
28	80.5	5.6	1237	2	A54080	protein-tyrosine-p
29	80	5.6	171	2	T01466	iron-sulfur cofact
30	80	5.6	280	2	JC7852	chitosanase (EC 3.
31	80	5.6	328	1	A41927	insulin-like growt
32	80	5.6	449	1	LIPG	triacylglycerol li
33	79.5	5.5	489	2	B70514	probable integral
34	79.5	5.5	910	2	S68983	auxilin - bovine
35	79.5	5.5	3432	1	GNWVJE	genome polyprotein
36	79.5	5.5	3432	1	GNWVJS	genome polyprotein
37	79.5	5.5	4367	1	B54802	dynein heavy chain
38	79	5.5	164	2	AC2218	hypothetical prote
39	79	5.5	525	2	S28052	sucrose transport
40	79	5.5	3507	2	T34513	hypothetical prote
41	78.5	5.5	398	2	A35281	integumentary muc
42	78.5	5.5	482	2	F71969	hypothetical prote
43	78.5	5.5	3083	2	AH2493	hypothetical prote
44	78.5	5.5	5376	2	T42215	zonadhesin - mouse
45	78.5	5.5	7962	2	I38346	elastic titin - hu

#### ALIGNMENTS

##### RESULT 1

S44605

C02F5.3 protein - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 02-Feb-2001

C;Accession: S44605

R;Anderson, K.

submitted to the EMBL Data Library, May 1993

A;Description: Sequence of the *C. elegans* cosmid C02F5.

A;Reference number: S44603

A;Accession: S44605

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-573 <AND>

A;Cross-references: EMBL:L14745; NID:g289607; PID:g289610

C;Genetics:

A;Introns: 224/2; 304/1; 363/3; 390/3; 503/2

C;Superfamily: translation elongation factor Tu homology

C;Keywords: GTP binding; nucleotide binding; P-loop

F;63-183/Domain: translation elongation factor Tu homology <ETU>

F;69-76/Region: nucleotide-binding motif A (P-loop)  
F;246-249/Region: GTP-binding NKXD motif

Query Match 11.6%; Score 167.5; DB 2; Length 573;  
Best Local Similarity 27.9%; Pred. No. 1.3e-06;  
Matches 50; Conservative 23; Mismatches 57; Indels 49; Gaps 5;

```
Qy      90 VTTGPWGAVATSAGGEESLKCEDLVGQYICKDP-----K 124
          |:| | | |      ::| | :   :|:| |
Db      415 VSTNPLGPV-----VECRFLENSFILCEDVPVLYGPGQTGQQPANESFRNEGKCLK 465

Qy      125 INDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYK 184
          :   | | | | | | | | | | | | | | | | | | | | : :
Db      466 MGGYRAEDVEFTN----VKCRVLPKIEC---HGPR-----FTKSTPCIIYNGHYFL 510

Qy      185 VAVALSFLGLGADRFYLGYPALGLLKFCTVGFCGIGSLIDFILISMQIVGPSDGSSY 243
          : |::| : | | | | | : : | | : | | : : :|:| | | :
Db      511 TTLLYSIFLGVVAVDRFCLGYSAMAVGKMLTLGGFGIWWIVDIFLLVLGVLGPADDSSW 569
```

## RESULT 2

T28787

hypothetical protein C41D11.5 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999

C;Accession: T28787

R;Gattung, S.; Maggi, L.

submitted to the EMBL Data Library, May 1997

A;Description: The sequence of *C. elegans* cosmid C41D11.

A;Reference number: Z20522

A;Accession: T28787

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-753 <GAT>

A;Cross-references: EMBL:AF003740; PIDN:AAC48141.1; GSPDB:GN00019; CESP:C41D11.5

A;Experimental source: strain Bristol N2; clone C41D11

C;Genetics:

A;Gene: CESP:C41D11.5

A;Map position: 1

A;Introns: 53/2; 81/3; 117/1; 250/3; 274/2; 357/3; 443/2; 485/3; 544/3; 585/3;  
637/2

Query Match 11.1%; Score 159.5; DB 2; Length 753;  
Best Local Similarity 28.2%; Pred. No. 8.6e-06;  
Matches 46; Conservative 27; Mismatches 61; Indels 29; Gaps 5;

```
Qy      104 GEESLKCE---DLKVGQYICKDPKINDATQE----PVNCTNYTA-----HVSCEF 145
          |   | |   | :| :   | :| : :   ||: | |   |
Db      284 GSAGLTCTFPGDCRIGDTV----KVNCTSRKGCPNPVSRNNVEAVCRFCWQLLPGDYDCE 339

Qy      146 PAPNITCKDS-----SGNETHFTGNEVGFFKPISCRNVNGYSYKVAVALSLFLGLWGA 198
          || | :   :   | : :   : | : | |   :||| :   ||: || ||
Db      340 PATNCSTSSTKLLVTKCSAHSSVICMGQRNFYKRIPCNWSSGYSWTKTMILSVVLGGFGA 399

Qy      199 DRFYLGYPALGLLKFCTVGFCGIGSLIDFILISMQIVGPSDGS 241
          ||||| : | : | | : :| :| :| : : : | |||
Db      400 DRFYLGWLKSAIGKLFSFGGLGVWTLVDVVLIAVGVIKPYDGS 442
```

RESULT 3

H75286

hypothetical protein - *Deinococcus radiodurans* (strain R1)

C;Species: *Deinococcus radiodurans*

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000

C;Accession: H75286

R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.;  
Dodson, R.J.; Haft, D.H.; Gwinn, M.L.; Nelson, W.C.; Richardson, D.L.; Moffat,  
K.S.; Qin, H.; Jiang, L.; Pamphile, W.; Crosby, M.; Shen, M.; Vamathevan, J.J.;  
Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Makarova, K.S.; Aravind, L.;  
Daly, M.J.; Minton, K.W.; Fleischmann, R.D.; Ketchum, K.A.; Nelson, K.E.;  
Salzberg, S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999

A;Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.

A;Reference number: A75250; MUID:20036896; PMID:10567266

A;Accession: H75286

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-309 <WHI>

A;Cross-references: GB:AE002064; GB:AE000513; NID:g6460134; PIDN:AAF11880.1;

PID:g6460145; TIGR:DR2326; GSPDB:GN00077

A;Experimental source: strain R1

C;Genetics:

A;Gene: DR2326

A;Map position: 1

Query Match 6.6%; Score 95; DB 2; Length 309;

Best Local Similarity 21.7%; Pred. No. 1.1;

Matches 73; Conservative 27; Mismatches 95; Indels 142; Gaps 16;

Qy	16	GQKNTRRDGTGLYPM---RGPFKNLALLPFS-LPLLGGGGSGSGEK---VSVSKMAAAW	67
		: :     :               :	
Db	11	GRLARQQRKGLDFRPVAEGERGPV-----FSPTPPFGGRNSGPPVRRVLSVMTDKDRDAG	63
Qy	68	PSG-----PSAPEAVTAR-----LVGVLF-----	87
Db	64	PSGNAPSWVDEVLSSSSSAPRPVEGRHGQTADPAQNPAGTAPGSGWDHWPQTDAARDLRL	123
Qy	88	-----VSVTTGPWGAVATSAGGEESLKCEDLVKGQYICKDPKINDATQEPVNCTNY	138
		:              :  :	
Db	124	PGDPPRPAPPSFDSDDWAARAT--GGE-----VRDPQGRD-----	156
Qy	139	TAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYK-----	184
		:    : :   :     :     :	
Db	157	-----PQESRTTVYSAAPQTDAGDPVRPAPPAPVKPVRGQMGSNGPAGLPVREDIA	209
Qy	185	----VAVALSLFLGWLGADRFYLGYPALGLLKF-CTVGF-----CGIGSLI----	225
		:    :       :        :  : ::	
Db	210	QKKLIAGLLGIFLGLGVHGFYLGQNGAGLLMLGVNIGVWVLAIVLSLLTLGLGAILFP	269
Qy	226	--DFILISMQIVGPSDGSSYII-----DY-YGTR	251
		: : :  :   :       :	
Db	270	LAGFVTSVLGVIGLIEGILYLTKSDADFQRDYLYGNK	306

# RESULT 4

S55661

hypothetical protein 66 - equine herpesvirus 2

C;Species: equine herpesvirus 2

C;Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 08-Oct-1999

C;Accession: S55661

R;Telford, E.A.R.; Watson, M.S.; Aird, H.C.; Perry, J.; Davison, A.J.

J. Mol. Biol. 249, 520-528, 1995

A;Title: The DNA sequence of equine herpesvirus 2.

A;Reference number: S55594; MUID:95302501; PMID:7783207

A;Accession: S55661

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-456 <TEL>

A;Cross-references: GB:U20824; NID:g695172; PIDN:AAC13854.1; PID:g695239

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995

Query Match 6.4%; Score 92; DB 2; Length 456;

Best Local Similarity 26.2%; Pred. No. 3.2;

Matches 42; Conservative 11; Mismatches 45; Indels 62; Gaps 7;

```

Qy      47 LGGGGSGSGEKVSVSKMAAAWPSGPSAPEAVTARLVGVLFVSVTTGPWGAVATSAGGEE 106
          ||||| | |:                               |:|                               |||
Db      13 LGGGGGGGGD-----LLG-----GGEA 29

Qy      107 S----LKCEDLKVGQYIC-KDPKINDATQEPVNCNTNYTAHV-----SCFPAPNITCKDS 155
          || |:|: | : | : | : ||: || | :|| |
Db      30 DGLMRALCEGLRVGEEDCARFVLYGVAYWQGGRCPEWVAHITRCADLSCFATYLLTCHRS 89

Qy      156 SGNETHFTGNEVGFFKPISCRNVNGYSYKVAVALSLFLGW 195
          | | ||| | : | | | | ||| :
Db      90 GGCE--FTGGRVARDRLPSLRE-----SVEVLQSLFLAF 121

```

# RESULT 5

VGIH59

E2 glycoprotein precursor - murine hepatitis virus (strain A59)

N;Alternate names: peplomer glycoprotein; spike glycoprotein

C;Species: murine hepatitis virus, MHV

C;Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 12-Apr-1996

C;Accession: A27402

R;Luytjes, W.; Sturman, L.S.; Bredenbeek, P.J.; Charite, J.; van der Zeijst, B.A.M.; Horzinek, M.C.; Spaan, W.J.M.

Virology 161, 479-487, 1987

A;Title: Primary structure of the glycoprotein E2 of coronavirus MHV-A59 and identification of the trypsin cleavage site.

A;Reference number: A27402; MUID:88072088; PMID:2825419

A;Accession: A27402

A;Molecule type: genomic RNA

A;Residues: 1-1324 <LUY>

C;Superfamily: coronavirus E2 glycoprotein

C;Keywords: glycoprotein; transmembrane protein

F;1-16/Domain: signal sequence #status predicted <SIG>

F;17-1324/Product: E2 glycoprotein #status predicted <E2G>

F;17-717/Product: 90B glycoprotein #status predicted <EGB>

F;718-1324/Product: 90A glycoprotein #status predicted <EGA>  
F;1266-1286/Domain: transmembrane #status predicted <TMN>  
F;31,60,192,247,357,435,442,530,625,657,665,688,737,754,844,893,1126,1180,1190,1209,1225,1246,1318/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 6.3%; Score 91; DB 1; Length 1324;  
Best Local Similarity 23.6%; Pred. No. 13;  
Matches 59; Conservative 27; Mismatches 94; Indels 70; Gaps 13;

```
Qy      25 TGLYPMRG-PFKNLALLP-----FSLPLLGGGGSGSGEKVSVSKMAAAWPSGPSA-- 73
      || ||: | |:|||| | | | | | : : ||| :|
Db      66 TGYYPVDGSKFRNLALTGTNSVSLSWFQPPYLNQFNDGIFAK--VQNLKTSTPSGATAYF 123

Qy      74 PEAVTARLVGVLFVSVTTGPGWAVATSAGGEESLKCEDLVKGQY-ICKDPKINDATQEP 132
      | | | | : :| | : | : | | | | : |
Db     124 PTIVIGSLFGYTSY-TVVIEPYNGVIMAS-----VCQYTICQLP----- 161

Qy     133 VNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNE-VGFF----KPISCRNVNGYSYKVAV 187
      : | | | ||: ||: | | | : : |
Db     162 -----YTDCKPNTN-----GNKLIGFWHTDVKPPICVLKRNFNLNVA 199

Qy     188 ALSLFLGWLGADRFYLGY---PALGLLKFC TVGFCGIGSLIDFILISMQIVGPSDGSSYI 244
      | : || | | : | || : : | | : ||:
Db     200 DAFYFHFYQHGGTFYAYYADKPSATTFLEFSVY----IGDILTQYYVLPFCNPTAGSTFA 255

Qy     245 IDYYGTRLTR 254
      | : | | :
Db     256 PRYWVTPLVK 265
```

RESULT 6

T08604

hypothetical protein GRR1 - soybean

C;Species: Glycine max (soybean)

C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 11-Jun-1999

C;Accession: T08604

R;Chen, W.; Atherly, A.

submitted to the EMBL Data Library, August 1997

A;Reference number: Z15438

A;Accession: T08604

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-690 <CHE>

A;Cross-references: EMBL:AF019910; NID:g2407789; PID:g2407790

A;Experimental source: variety L85-3044; root

C;Genetics:

A;Gene: grr1

Query Match 6.2%; Score 89.5; DB 2; Length 690;  
Best Local Similarity 24.6%; Pred. No. 8.4;  
Matches 43; Conservative 22; Mismatches 59; Indels 51; Gaps 7;

```
Qy      85 LWFVSVTTGPGWAVATSAGGEESLKCEDLVKGQYICKDPKINDATQEPV--NCTNYTAHV 142
      || |: | | | : :| : | : :|| | |:| | | || | | :
Db     233 LWDVA-TVGDVGLIEIASGCHQLEKLD-----LCKCPNISDKTLIAVAKNCPN-LAEL 283

Qy     143 SCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAVALSLFLGWLGADRFY 202
```



```

      |   |||   |||   :   : || :| :|   :
Db      284 SIЕСPNI-----GNEGLQAIGKCPNLRISISIKNCSGVGDQ----- 319

Qy      203 LGYPALGLLKFCTVGFSGIGSLIDFIL--ISMQIVGPSDGSSYIIDYYGTRLTRL 255
      | | : | | | : : : | | | : | : | | : | |
Db      320 -----GVAGLLSSASFALTKVKLES LTVSDLSLAVIGHYGVAVTDL 360

```

# RESULT 7

H75632

Na(+)-linked D-alanine glycine permease - *Deinococcus radiodurans* (strain R1)

C;Species: *Deinococcus radiodurans*

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000

C;Accession: H75632

R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.;  
 Dodson, R.J.; Haft, D.H.; Gwinn, M.L.; Nelson, W.C.; Richardson, D.L.; Moffat,  
 K.S.; Qin, H.; Jiang, L.; Pamphile, W.; Crosby, M.; Shen, M.; Vamathevan, J.J.;  
 Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Makarova, K.S.; Aravind, L.;  
 Daly, M.J.; Minton, K.W.; Fleischmann, R.D.; Ketchum, K.A.; Nelson, K.E.;  
 Salzberg, S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999

A;Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.

A;Reference number: A75250; MUID:20036896; PMID:10567266

A;Accession: H75632

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-547 <WHI>

A;Cross-references: GB:AE001826; NID:g6460827; PIDN:AAF12563.1; PID:g6460859;  
 TIGR:DRB0133; GSPDB:GN00079

A;Experimental source: strain R1

C;Genetics:

A;Gene: DRB0133

A;Map position: megaplasmid

A;Genome: plasmid

A;Note: plasmid MP1

C;Superfamily: sodium-dependent D-alanine/glycine transport protein

```

Query Match          6.1%; Score 87.5; DB 2; Length 547;
Best Local Similarity 25.6%; Pred. No. 9.6;
Matches 57; Conservative 19; Mismatches 52; Indels 95; Gaps 13;

```

```

Qy      43 SLPLLGGGGSGSGEKVSVSKMAAA-----WPS--GPSAPEAVTARLVGV----- 84
      | ||   || | : : : ||           | | ||| | : | : |
Db      29 SRPLSSESGSSSAQEPWMGRLPAALVFTGLLGAVSWASAQGPSVDERINAVVTPVSHFLS 88

Qy      85 -LWFSVVTGP-----WGAVATSAGGEESLKCE-----DLKVGQYIC 120
      | | | : |   | |||   : |   | | | : |
Db      89 GLIFASISVGEAQVPLIVVWLAVA-----AIVCTLSFRFVNIWGFKHGIDLVRGRY-- 139

Qy     121 KDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNG 180
      |||           | : | |||   : | : |
Db     140 ----GNDA-----DAPGMVTHFQALTAVSGTVGLGNIAG 170

Qy     181 YSYKVAVALSLFLGWLGADRFYLGYPALGLL----KF--CTVG 217
      ||||| | | | : : : ||| | | : |
Db     171 ----VAVALS--LGGPGATFWMI---LVGLLSMSTKFVECTLG 204

```

## T23754

C;Species: *Caenorhabditis elegans*

C>Date: 15-Oct-1999 #sequence revision 15-Oct-1999 #text change 29-Oct-1999

C;Accession: T23754; T24513

R; Thomas, K.

submitted to the EMBL Data Library, June 1995

A;Reference number: Z19793

A;Accession: T23754

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-1207 <WIL>

A:Cross-references: EMBL:Z49968; PIDN:CAA90265.1; GSPDB:GN00020; CESP:T05C12.10

A;Experimental source: clone M110

R; Burton, J.

submitted to the EMBL Data Library, October 1995

A;Reference number: Z19901

A;Accession: T24513

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-1207 <WI2>

A;Cross-references: EMBL:Z66500; PIDN:CAA91313.1; GSPDB:GN00020; CESP:T05C12.10

A;Experimental source: clone T05C12

C; Genetics:

A;Gene: CESP:T05C12.10

```
A;Map position: 2
```

A; Introns: 31/3; 87/2; 141/3; 180/2; 203/3; 267/1; 776/2; 794/2; 834/2; 1086/3;

1143/1; 1181/1

Query Match 6.1%; Score 87.5; DB 2; Length 1207;

Best Local Similarity 21.1%; Pred. No. 23;

Matches 47; Conservative 26; Mismatches 83; Indels 67; Gaps 8;

Qy            2 HILKGSNPVIPRAHGQKNTRRDGTGLYPMRGPFKNLALLPFSLP LLGGGSGSGEKVSVS    61

| : : : | : | : | : | : | : | : | : | : | : | : | :

Db 848 HNAESSASGIPLVOARSNTVNGGAPVPPAPGS-----GATGSGTSGSGTSESVT 896

Qv 62 KMAAAWPSGPS-----APEAVTARLVGVLFVSVTTGPWGAVATSAG 103

Db 897 NGSGATESGSGTGTGTSGTGSSGTGASAARTSSIAGDAPQAAVLADTPGAAGAAGG 956

QY 104 GE-----ESL-----KCEDLVGQYICKDPKINDAT-----QEPVNCTNYTA 140  
| :|| : :||| : :| :|| ||:

Db 957 GRSNCF SADSLVTTVTGQKRMDELQIGDYVLVPSSGNVLKYEKVEMFYHREP KTRTNF-- 1014

Ov 141 HVSCFPAPNITCKDSSGNETHFTGNEVGFFKPIS-CRNVNGYS 182

$$I_1 \quad : \quad || \quad : \quad || \quad : \quad | :: \quad | \quad | \quad | :$$

Db 1015 -----VVLTKSGRKLSLTGRHL---LPVAECSQVEQYT 1045

## RESULT 9

S20911

alcohol dehydrogenase (EC 1.1.1.1) II - yeast (*Kluyveromyces marxianus* var. *lactis*)

C;Species: Kluyveromyces marxianus var. lactis, Candida sphaerica  
 C;Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
 C;Accession: S20911; S19804  
 R;Shain, D.H.; Salvadore, C.; Denis, C.L.  
 Mol. Gen. Genet. 232, 479-488, 1992  
 A;Title: Evolution of the alcohol dehydrogenase (ADH) genes in yeast:  
 characterization of a fourth ADH in Kluyveromyces lactis.  
 A;Reference number: S20911; MUID:92269769; PMID:1588917  
 A;Accession: S20911  
 A;Molecule type: DNA  
 A;Residues: 1-348 <SHA>  
 A;Cross-references: EMBL:X64397; NID:g2832; PIDN:CAA45739.1; PID:g2833  
 C;Genetics:  
 A;Gene: ADH2  
 C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology  
 C;Keywords: alcohol metabolism; metalloprotein; NAD; oxidoreductase; zinc  
 F;29-336/Domain: long-chain alcohol dehydrogenase homology <LADH>  
 F;173-202/Region: beta-alpha-beta NAD nucleotide-binding fold  
 F;44,67,154/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted  
 F;98,101,104,112/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match 5.9%; Score 85.5; DB 1; Length 348;  
 Best Local Similarity 19.9%; Pred. No. 8.5;  
 Matches 65; Conservative 35; Mismatches 111; Indels 115; Gaps 13;

```

QY      19 NTRRDG---TGLYPMRGPFKNLALLPFSPLLLGGGGSGSGEKVSVSKMAAAW----- 67
      | : | | : : | | | | | : | | : : |
Db      37 NVKYSQVCHTDLHAWKGDWP----LPTKPLPV-GGHEGAGVVVAMGENVKGWIIGDFAGI 91

QY      68 -----PSGPSAPEAVT 78
      | | |
Db      92 KWLNGSCMSCEYCELSNESNCPDADLSGYTHDGSFQQYATADAVQAARIPKGTDLAEVAP 151

QY      79 ARLVGV-----LWFVSVTTGPGAVATSAGGEESLKCEDLKV----- 115
      || | : | | | : | | | : |
Db     152 ILCAGVTVYKALKSADLKAGDWVAISGACGGLGSLAIQYAKAMGYRVLGIDTGAEKAKLF 211

QY     116 ----GQYICKDPKINDATQEPVNCTNYTAH-----VSCFPAPNITCKDSSGNETHFTGN 165
      | : | | : | | | | | | | : | | |
Db     212 KELGGEYFVDYAVSKDLIKEIVDATNGGAHGVINVSVEFAI-----EQSTNYVRSNGT 265

QY     166 EVGFFKPISCRNVNGYSYKVAVALSLFLGWLG--AD-RFYLGYPALGLLKFCCTVGFCGIG 222
      | | : : : | : : : : | | | | : | | : | :
Db     266 VVLVGLPRDAKCKSDVFTQVVKSVSIVGSYVGNRADTREALDFFARGLV-HAPIKIVGLS 324

QY     223 SLIDFI--LISMQIVGPSDGSSSYIID 246
      | | : : : | | : |
Db     325 ELADVYDKMVKGGEIVG-----RYVVD 345

```

# RESULT 10

T35005

probable integral membrane transporter - Streptomyces coelicolor

C;Species: Streptomyces coelicolor

C;Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 17-Mar-2000

C;Accession: T35005

R;Seeger, K.J.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.;  
 Rajandream, M.A.  
 submitted to the EMBL Data Library, December 1998  
 A;Reference number: Z21564  
 A;Accession: T35005  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-446 <SEE>  
 A;Cross-references: EMBL:AL034443; PIDN:CAA22367.1; GSPDB:GN00070;  
 SCOEDB:SC4B5.13  
 A;Experimental source: strain A3(2)  
 C;Genetics:  
 A;Gene: SCOEDB:SC4B5.13  
 C;Superfamily: hypothetical protein c0103

Query Match 5.8%; Score 84; DB 2; Length 446;  
 Best Local Similarity 19.7%; Pred. No. 15;  
 Matches 52; Conservative 34; Mismatches 110; Indels 68; Gaps 9;

```

Qy      39 LLPFSLPLLGGGGSGSGEKVSVSKMAAAWPSGPSAPEAVTARLVGVLF--VSVTTGPWG 96
      | :      : || | | |      || :      : | |      | : :: |      :: ||
Db      106 LIVLGYGFVGGIGLIGYISPVSTLIKWFDPDPG-----MATGIAIMGFGGGALIASPWS 160

Qy      97 AVATSAGGEES-----LKCEDLKVGQYICKDPK-----IN 126
      |      : | ::      : : |      : : | :
Db      161 AQMLKSFGTDNSGIALAFLVHGLTYAVFMLLGVLLVRVPRPRERADGRPAPLEGVQVSAR 220

Qy     127 DATQEP-----VNCTNYTAHVSCF--PAPNITCKDSSGNETHFTGNEVGFFKPISCRN 177
      | : |      | | | | :      | | | |      | :      | | : | |
Db     221 SAVRTPQFWLLWIVLCMNVTAGIGILEKAAPMITDFFSDTSTPVSVTAAAGFVALLSAAN 280

Qy     178 VNGYSYKVAVALSLFLGW-----LGADRFYLGYPALGLLKFCTVGFCGIGSLIDFILIS 231
      : |      ||      : |      | | : | | : :      | |      | : | :
Db     281 MAGR-----FGWSSASDLIGRKNIYRVYLGVGALMYTLIALFGDSSKPLFVLCA 329

Qy     232 MQIV-----GPSDGSSYIIDYYGT 250
      : : |      | :      : | : | : | |
Db     330 LVVVSFYGGGFATAPAYLKDLFGT 353
  
```

# RESULT 11

C70574

probable aroP2 protein - Mycobacterium tuberculosis (strain H37RV)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C;Accession: C70574

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.;  
 Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry III, C.E.; Tekaiia, F.; Badcock, K.;  
 Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;  
 Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.;  
 Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail,  
 M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.;  
 Squares, S.

Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete  
 genome sequence.

A;Reference number: A70500; MUID:98295987; PMID:9634230  
A;Accession: C70574  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-487 <COL>  
A;Cross-references: GB:Z95324; GB:AL123456; NID:g3261760; PIDN:CAB08578.1;  
PID:g2094825  
A;Experimental source: strain H37Rv  
C;Genetics:  
A;Gene: aroP2  
C;Superfamily: arginine permease

Query Match 5.8%; Score 83.5; DB 2; Length 487;  
Best Local Similarity 26.5%; Pred. No. 19;  
Matches 27; Conservative 16; Mismatches 40; Indels 19; Gaps 5;

```
Qy      159 ETHFTGNEVGFFKPISCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCTVGF 218
      : | : | : | : : : | : | | | | | | | | |
Db      8  DERLTREDTGYHKGLHSRQLQMIALGGAIGTGLFLG--AGGRLASAGPGL----FLVYGI 61

Qy      219 CGIGSLIDFILISMQIVG-----PSDGS--SYIIDYYGTRL 252
      | | | | : : : : | | | | | : | : |
Db      62  CGI-----FVFLILRALGELVLHRPSSGSFVSYAREFYGEKV 98
```

#### RESULT 12

B75447

hypothetical protein - *Deinococcus radiodurans* (strain R1)

C;Species: *Deinococcus radiodurans*

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000

C;Accession: B75447

R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.;  
Dodson, R.J.; Haft, D.H.; Gwinn, M.L.; Nelson, W.C.; Richardson, D.L.; Moffat,  
K.S.; Qin, H.; Jiang, L.; Pamphile, W.; Crosby, M.; Shen, M.; Vamathevan, J.J.;  
Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Makarova, K.S.; Aravind, L.;  
Daly, M.J.; Minton, K.W.; Fleischmann, R.D.; Ketchum, K.A.; Nelson, K.E.;  
Salzberg, S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999

A;Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans*  
R1.

A;Reference number: A75250; MUID:20036896; PMID:10567266

A;Accession: B75447

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-137 <WHI>

A;Cross-references: GB:AE001954; GB:AE000513; NID:g6458751; PIDN:AAF10608.1;  
PID:g6458763; TIGR:DR1033; GSPDB:GN00077

A;Experimental source: strain R1

C;Genetics:

A;Gene: DR1033

A;Map position: 1

Query Match 5.8%; Score 83; DB 2; Length 137;  
Best Local Similarity 26.5%; Pred. No. 4.9;  
Matches 39; Conservative 16; Mismatches 48; Indels 44; Gaps 8;

```
Qy      33 PFKNL---ALLPFSLPLLGGGGSGSGEKVSVSKMAAAWPSGPSAP-EAVTARLVGVLFV 88
```

```

      | | |   | |   |   | | |   |   : | : | |   | |   : | |   | |
Db      17 PIKKLLPVVLLASVLTACGGGTSTPG-----TSTPNTPAVPSSAVAPKLSG---FV 64

Qy      89 -----SVT---TGPWGAVATSAGGEESLKCEDLKVGQYICKDPKINDATQEPVNCTNYT 139
      | : |   |   |   | | |   : :   | :   |   |
Db      65 LSGSQHSLTVSLNAPASCVFNSAAGSLNMTAATLEGSPYA-----YA 106

Qy      140 AHVS-CFPAPNITCKDSSGNETHFTGN 165
      : | : | : : | | : : | : : |   | |
Db      107 VSLSGSYPKASVTCTNSAGSDTLNLGN 133

```

RESULT 13

S32521

alcohol dehydrogenase (EC 1.1.1.1) 1 - yeast (*Kluyveromyces marxianus* var. *marxianus*)

C;Species: *Kluyveromyces marxianus* var. *marxianus*, *Candida kefyr*

C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 28-Jul-2000

C;Accession: S32521

R;Ladriere, J.M.; Delcour, J.; Vandenhoute, J.

Biochim. Biophys. Acta 1173, 99-101, 1993

A;Title: Sequence of a gene coding for a cytoplasmic alcohol dehydrogenase from *Kluyveromyces marxianus* ATCC 12424.

A;Reference number: S32521; MUID:93250057; PMID:8485163

A;Accession: S32521

A;Molecule type: DNA

A;Residues: 1-348 <LAD>

A;Cross-references: EMBL:X60224; NID:g6822201; PIDN:CAA42785.1; PID:g297908

C;Genetics:

A;Gene: ADH1

C;Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

C;Keywords: alcohol metabolism; metalloprotein; NAD; oxidoreductase; zinc

F;29-336/Domain: long-chain alcohol dehydrogenase homology <LADH>

F;44,67,154/Binding site: zinc, catalytic (Cys, His, Cys) #status predicted

F;98,101,104,112/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match 5.8%; Score 83; DB 1; Length 348;

Best Local Similarity 20.8%; Pred. No. 14;

Matches 65; Conservative 38; Mismatches 107; Indels 102; Gaps 14;

```

Qy      19 NTRRDG---TGLYPMRGPFKNLALLPFSLLPGGGSGSGEKVSVSKMAAAWPSG----- 70
      | : |   | | : : |   |   | | | : | | : :   | |
Db      37 NVKYSGVCHTDLHAWQGDWP----LDTKLPLV-GGHEGAGIVVAMGENVTGWEIGDYAGI 91

Qy      71 -----PSAPEA-----VTARLV----- 82
      | : | : |   | | |
Db      92 KWLNGSCMSCEECELSNEPNCPKADLSGYTHDGSFQQYATADAVQAARIPKNVDLAEVAP 151

Qy      83 ----GV-----LWEFVSVTTGPGWAVATSAGGEESLKCEDLV----- 115
      | |   |   : | | | : : | | | : |
Db      152 ILCAGVTVYKALKSAHIKAGDWVAISGACGGLGSLAIQYAKAMGYRVLGIDAGDEKAKLF 211

Qy      116 ----GQYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNE--VGF 169
      | : |   | | | : | | | : :   : | | | |
Db      212 KELGGEYFIDFTKTKDMVAEVIEATNGVAHAVINVSVSEAAISTSVLYTRNGTVVLVGL 271

Qy      170 FKPISCRNVNGYSYKVAVALSLFLGWLG--AD-RFYLGYPALGLLK--FCTVGFCGIGSL 224

```

```

      : | : : | ::| ::| || | | : : ||:| : | : | :
Db      272 PRDAQCK--SDVFNQVVKSISIVGSYVGNRADTREALDFFSRGLVKAPIKILGLSELASV 329

Qy      225 IDFILISMQIVG 236
      | : : ||||
Db      330 YD-KMVKGQIVG 340

```

# RESULT 14

BBHU

complement factor B precursor [validated] - human

N;Alternate names: C3 convertase; C3 proactivator; glycine-rich beta-glycoprotein; heat-labile complement factor; proenzyme factor B; properdin factor B

N;Contains: alternative-complement-pathway C3/C5 convertase (EC 3.4.21.47) Bb fragment

C;Species: Homo sapiens (man)

C;Date: 19-Feb-1984 #sequence\_revision 05-Aug-1994 #text\_change 08-Dec-2000

C;Accession: S34075; A44622; A00934; A19188; A19947; B19947; B25971; S14339; A44628; I54409; I57824; B19447

R;Mejia, J.E.; Jahn, I.; de la Salle, H.; Hauptmann, G.

submitted to the EMBL Data Library, March 1993

A;Reference number: S34075

A;Accession: S34075

A;Molecule type: mRNA

A;Residues: 1-764 <MEJ>

A;Cross-references: EMBL:X72875; NID:g297568; PIDN:CAA51389.1; PID:g297569

R;Woods, D.E.; Markham, A.F.; Ricker, A.T.; Goldberger, G.; Colten, H.R.

Proc. Natl. Acad. Sci. U.S.A. 79, 5661-5665, 1982

A;Title: Isolation of cDNA clones for the human complement protein factor B, a class III major histocompatibility complex gene product.

A;Reference number: A44622; MUID:83039428; PMID:6957884

A;Accession: A44622

A;Molecule type: mRNA

A;Residues: 467-546;550-595;752-764 <WOO>

A;Cross-references: GB:J00185; GB:J00186

A;Note: the authors translated the codon TAC at 519 as Thr; the nucleic acid translation differs from the sequence shown in having 537-Thr, and 764-His

R;Mole, J.E.; Anderson, J.K.; Davison, E.A.; Woods, D.E.

J. Biol. Chem. 259, 3407-3412, 1984

A;Title: Complete primary structure for the zymogen of human complement factor B.

A;Reference number: A20751; MUID:84161997; PMID:6546754

A;Accession: A00934

A;Molecule type: protein; mRNA

A;Residues: 26-764 <MOL>

A;Cross-references: GB:K01566

A;Note: nucleic acid translation differs from the sequence shown in having 300-Leu, 328-Val, 356-Glu, and 357-Glu

A;Note: 736-Ser was also found

A;Note: glycosylation sites were determined

R;Christie, D.L.; Gagnon, J.

Biochem. J. 209, 61-70, 1983

A;Title: Amino acid sequence of the Bb fragment from complement factor B.

Sequence of the major cyanogen bromide-cleavage peptide (CB-II) and completion of the sequence of the Bb fragment.

A;Reference number: A19188; MUID:83204002; PMID:6342610

A;Contents: the final paper in a series documenting the sequence, glycosylation site, and active site  
 A;Accession: A19188  
 A;Molecule type: protein  
 A;Residues: 260-296,'T',298-764 <CHR>  
 R;Campbell, R.D.; Porter, R.R.  
 Proc. Natl. Acad. Sci. U.S.A. 80, 4464-4468, 1983  
 A;Title: Molecular cloning and characterization of the gene coding for human complement protein factor B.  
 A;Reference number: A19947; MUID:83273641; PMID:6308626  
 A;Accession: A19947  
 A;Molecule type: DNA  
 A;Residues: 346-764 <CAM>  
 A;Cross-references: GB:J00125  
 A;Accession: B19947  
 A;Molecule type: mRNA  
 A;Residues: 339-509 <CA1>  
 A;Cross-references: GB:J00126; NID:g187723; PIDN:AAA36226.1; PID:g553536  
 R;Wu, L.; Morley, B.J.; Campbell, R.D.  
 Cell 48, 331-342, 1987  
 A;Title: Cell-specific expression of the human complement protein factor B gene: evidence for the role of two distinct 5'-flanking elements.  
 A;Reference number: A25971; MUID:87102880; PMID:3643061  
 A;Accession: B25971  
 A;Molecule type: DNA  
 A;Residues: 1-99 <WUL>  
 A;Cross-references: GB:M15082; NID:g187699; PIDN:AAA59625.1; PID:g553534  
 R;Niemann, M.A.; Bhowan, A.S.; Miller, E.J.  
 Biochem. J. 274, 473-480, 1991  
 A;Title: The principal site of glycation of human complement Factor B.  
 A;Reference number: S14339; MUID:91174758; PMID:2006911  
 A;Accession: S14339  
 A;Molecule type: protein  
 A;Residues: 270-329 <NIE>  
 A;Note: binding site for carbohydrate to lysine under artificial conditions  
 R;Morley, B.J.; Campbell, R.D.  
 EMBO J. 3, 153-157, 1984  
 A;Title: Internal homologies of the Ba fragment from human complement component factor B, a class III MHC antigen.  
 A;Reference number: A44628; MUID:84158524; PMID:6323161  
 A;Accession: A44628  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 16-225,'F',227-259 <MOR>  
 R;Schwaeble, W.; Luttig, B.; Sokolowski, T.; Estaller, C.; Weiss, E.H.; Meyer zum Buschenfelde, K.H.; Whaley, K.; Dippold, W.  
 Immunobiology 188, 221-232, 1993  
 A;Title: Human complement factor B: functional properties of a recombinant zymogen of the alternative activation pathway convertase.  
 A;Reference number: I54409; MUID:94041399; PMID:8225386  
 A;Accession: I54409  
 A;Status: translated from GB/EMBL/DDBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-764 <RES>  
 A;Cross-references: GB:S67310; NID:g452937; PIDN:AAD13989.1; PID:g4261689  
 R;Horiuchi, T.; Kim, S.; Matsumoto, M.; Watanabe, I.; Fujita, S.; Volanakis, J.E.



Mol. Immunol. 30, 1587-1592, 1993

A;Title: Human complement factor B: cDNA cloning, nucleotide sequencing, phenotypic conversion by site-directed mutagenesis and expression.

A;Reference number: I57824; MUID:94067177; PMID:8247029

A;Accession: I57824

A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-31,'Q',33-764 <RE2>

A;Cross-references: GB:L15702; NID:g291921; PIDN:AAA16820.1; PID:g291922

C;Comment: 292-Cys has a free sulfhydryl.

C;Genetics:

A;Gene: GDB:BF

A;Cross-references: GDB:119726; OMIM:138470

A;Map position: 6p21.3-6p21.3

A;Introns: 21/3; 99/3; 346/1; 390/1; 424/1; 470/1; 502/3; 542/1; 593/2; 619/1; 652/3; 697/1; 713/3

A;Note: the list of introns may be incomplete

A;Note: gene is located in the major histocompatibility complex, class III region

C;Complex: complement factor B initially forms an inactive complex with complement factor C3b, becoming susceptible to cleavage by factor D into Ba and Bb fragments; Bb remains associated with complement factor C3b forming active C3/C5 convertase; Ba is released

C;Function:

A;Description: Bb is a serine proteinase; C3/C5 convertase cleaves complement C3 alpha chain to release C3a and form C3b; it also cleaves C5 alpha chain to release C5a and form C5b; Ba is nonfunctional

A;Pathway: complement alternate pathway

C;Superfamily: complement C2; complement factor H repeat homology; trypsin homology; von Willebrand factor type A repeat homology

C;Keywords: acute phase; complement alternate pathway; duplication; glycoprotein; hydrolase; plasma; serine proteinase

F;1-25/Domain: signal sequence #status predicted <SIG>

F;26-764/Product: complement factor B #status experimental <MAT>

F;26-259/Product: complement factor Ba fragment #status experimental <BAF>

F;37-98/Domain: complement factor H repeat homology <FH1>

F;103-158/Domain: complement factor H repeat homology <FH2>

F;165-218/Domain: complement factor H repeat homology <FH3>

F;260-764/Product: C3/C5 convertase Bb fragment #status experimental <BBF>

F;268-458/Domain: von Willebrand factor type A repeat homology <VFA>

F;482-752/Domain: trypsin homology #status atypical <TRY>

F;37-76,62-98,103-145,131-158,165-205,191-218,478-596,511-527,599-615,656-682,695-725/Disulfide bonds: #status predicted

F;122,142,285,378/Binding site: carbohydrate (Asn) (covalent) #status experimental

F;259-260/Cleavage site: Arg-Lys (complement factor D) #status experimental

F;526,576,699/Active site: His, Asp, Ser #status experimental

Query Match 5.8%; Score 83; DB 1; Length 764;

Best Local Similarity 24.1%; Pred. No. 34;

Matches 49; Conservative 21; Mismatches 71; Indels 62; Gaps 12;

```
Qy      24 GTGLYPMRGPFKNLALLPFSLLPLGGG-----GSGSGEKVSV----- 60
          | : | |      | | : | | | | |      | | | | | :
Db      2 GSNLSP-----QLCLMPFILGLLSGGVTTTPWSLARPPQSGSCSLEGVEIKGGSFRLLQEG 55

Qy      61 SKMAAAWPSG--PSAPEAVTARLVGVLFVSVTTGPWGAVATS-----AGGEESLKC-- 110
```

```

      :   ||| |   :   | |   : || |   : |   : : |
Db      56 QALEYVCPSGFYPPVQTRTCR-----STGSWSTLKTQDQKTVRKAECRAIHCPR 105

Qy      111 -EDLKVGQYICKDPKINDATQEPVNC-TNYTAHVSCFPAPNITCKDSS--GNETHFTGNE 166
      | : | : | : | | : : : | | | | | | | : : : | | |
Db      106 PHDFENGEYWPRSPYYNVSEISFHCYDGYTLRGSA----NRTCQVNGRWSGQTAICDNG 161

Qy      167 VGFFK----PISCRNVNGYSYKV 185
      | :   || | | | | : :
Db      162 AGYCSNPGIPIGTRKV-GSQYRL 183

```

RESULT 15

T10729

transferrin-like protein Ttf-1, salt-induced - green alga (Dunaliella salina)

C;Species: Dunaliella salina

C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 16-Jul-1999

C;Accession: T10729

R;Fisher, M.; Gokhman, I.; Pick, U.; Zamir, A.

submitted to the EMBL Data Library, November 1996

A;Reference number: Z17101

A;Accession: T10729

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-1274 <FIS>

A;Cross-references: EMBL:U77059; NID:g1684791; PID:g1684792

C;Genetics:

A;Gene: ttf1

C;Superfamily: transferrin repeat homology

Query Match 5.8%; Score 83; DB 2; Length 1274;

Best Local Similarity 22.9%; Pred. No. 61;

Matches 30; Conservative 17; Mismatches 50; Indels 34; Gaps 6;

```

Qy      57 KVSVSKMAAAWPSGPSAPEAV-TARLVGVLFVSVTTGPGWAVATSAGGEESLKCEDLKV 115
      :|   :   |   || :   |:|   | :| | : ||
Db      579 QVDAETIEKFWEDNVCAPGSTENGPLIG-----GGKYGEVGENGGG----- 619

Qy      116 GQYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISC 175
      :|| | : :::| | | | | | | ||: ||| :: | | :
Db      620 ---LCKRCKTDCTSEDPY--AGYDGAHVCI-----DDDDGNQ--FTGGDIAFVKHSTL 665

Qy      176 RNVNGYSYKVA 186
      | : || : |
Db      666 RDYNGPNLNTA 676

```

Search completed: March 4, 2004, 10:27:47

Job time : 46 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2004, 10:26:56 ; Search time 652 Seconds  
(without alignments)  
87.117 Million cell updates/sec

Title: US-09-852-100B-2  
Perfect score: 1439  
Sequence: 1 MHILKGSPNVIPRAHGQKNT.....TRLTRLSITNETFRKTQLYP 269

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	Match	Length	DB	ID	Description
No.							

1	1439	100.0	269	9	US-09-852-100A-2	Sequence 2, Appli
2	1439	100.0	269	9	US-09-833-503A-2	Sequence 2, Appli
3	1439	100.0	269	14	US-10-199-881-2	Sequence 2, Appli
4	201	14.0	221	10	US-09-974-879-230	Sequence 230, App
5	201	14.0	222	10	US-09-305-736-230	Sequence 230, App
6	201	14.0	222	11	US-09-818-683-230	Sequence 230, App
7	200	13.9	221	9	US-09-833-503A-6	Sequence 6, Appli
8	200	13.9	221	10	US-09-992-600A-82	Sequence 82, Appl
9	200	13.9	221	10	US-09-924-340-82	Sequence 82, Appl
10	200	13.9	221	10	US-09-992-095B-82	Sequence 82, Appl
11	200	13.9	221	10	US-09-999-570-82	Sequence 82, Appl
12	200	13.9	221	14	US-10-000-489-82	Sequence 82, Appl
13	200	13.9	221	14	US-10-000-986-82	Sequence 82, Appl
14	200	13.9	221	14	US-10-199-881-6	Sequence 6, Appli
15	200	13.9	221	14	US-10-154-678-82	Sequence 82, Appl
16	200	13.9	247	10	US-09-796-753-48	Sequence 48, Appl
17	196	13.6	230	10	US-09-796-753-50	Sequence 50, Appl
18	167.5	11.6	573	15	US-10-369-493-5797	Sequence 5797, Ap
19	136.5	9.5	214	9	US-09-833-503A-4	Sequence 4, Appli
20	136.5	9.5	214	14	US-10-199-881-4	Sequence 4, Appli
21	136.5	9.5	224	9	US-09-925-301-1018	Sequence 1018, Ap
22	121	8.4	171	15	US-10-264-237-2752	Sequence 2752, Ap
23	121	8.4	172	10	US-09-798-889-58	Sequence 58, Appl
24	115.5	8.0	213	10	US-09-866-050A-631	Sequence 631, App
25	94.5	6.6	166	14	US-10-156-761-10938	Sequence 10938, A
26	92.5	6.4	87	14	US-10-156-761-10936	Sequence 10936, A
27	90	6.3	443	14	US-10-156-761-9371	Sequence 9371, Ap
28	87.5	6.1	939	15	US-10-369-493-3816	Sequence 3816, Ap
29	87	6.0	476	14	US-10-156-761-8693	Sequence 8693, Ap
30	85	5.9	338	9	US-09-808-602-72	Sequence 72, Appl
31	85	5.9	338	10	US-09-800-198-61	Sequence 61, Appl
32	84.5	5.9	389	9	US-09-925-297-489	Sequence 489, App
33	84.5	5.9	389	15	US-10-264-049-2190	Sequence 2190, Ap
34	84.5	5.9	718	15	US-10-108-260A-3375	Sequence 3375, Ap
35	83.5	5.8	179	14	US-10-156-761-13418	Sequence 13418, A
36	83	5.8	764	14	US-10-177-293-27	Sequence 27, Appl
37	83	5.8	764	15	US-10-236-031B-44	Sequence 44, Appl
38	82.5	5.7	338	9	US-09-808-602-69	Sequence 69, Appl
39	82.5	5.7	338	10	US-09-800-198-58	Sequence 58, Appl
40	82.5	5.7	338	14	US-10-161-572-58	Sequence 58, Appl
41	82.5	5.7	338	16	US-10-038-854-158	Sequence 158, App
42	82.5	5.7	5374	15	US-10-028-248A-75	Sequence 75, Appl
43	82.5	5.7	5374	15	US-10-107-782-75	Sequence 75, Appl
44	82	5.7	416	14	US-10-156-761-10376	Sequence 10376, A
45	81	5.6	717	14	US-10-054-044A-2	Sequence 2, Appli

#### ALIGNMENTS

RESULT 1  
 US-09-852-100A-2  
 ; Sequence 2, Application US/09852100A  
 ; Patent No. US20020058267A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: American Home Products

```

; TITLE OF INVENTION: Beta-amyloid Peptide-Binding Proteins and Polynucleotides
Encoding the
; TITLE OF INVENTION: Same
; FILE REFERENCE: AHP981261p2
; CURRENT APPLICATION NUMBER: US/09/852,100A
; CURRENT FILING DATE: 2001-05-09
; PRIOR APPLICATION NUMBER: US 09/172,990
; PRIOR FILING DATE: 1998-10-14
; PRIOR APPLICATION NUMBER: US 60/104,104
; PRIOR FILING DATE: 1998-10-13
; PRIOR APPLICATION NUMBER: PTC/US99/21621
; PRIOR FILING DATE: 1999-10-13
; PRIOR APPLICATION NUMBER: US 09/060,609
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: US 60/064,583
; PRIOR FILING DATE: 1997-04-16
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-852-100A-2

```

```

Query Match          100.0%; Score 1439; DB 9; Length 269;
Best Local Similarity 100.0%; Pred. No. 3e-135;
Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 MHILKGSPNVIPRAHGQKNTRRDGTGLYPMRGPFKNLALLPFSLPLLGGGGSGSGEKVSV 60
        |||
Db      1 MHILKGSPNVIPRAHGQKNTRRDGTGLYPMRGPFKNLALLPFSLPLLGGGGSGSGEKVSV 60

Qy     61 SKMAAAWPSGPSAPEAVTARLVGVLFVSVTTGPGWAVATSAGGEESLKCEDLKVGQYIC 120
        |||
Db     61 SKMAAAWPSGPSAPEAVTARLVGVLFVSVTTGPGWAVATSAGGEESLKCEDLKVGQYIC 120

Qy    121 KDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNG 180
        |||
Db    121 KDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNG 180

Qy    181 YSYKVAVALSLFLGWLGAADRFYLGYPALGLLKFCTVGFCGIGSLIDFILISMQIVGPSDG 240
        |||
Db    181 YSYKVAVALSLFLGWLGAADRFYLGYPALGLLKFCTVGFCGIGSLIDFILISMQIVGPSDG 240

Qy    241 SSIIDYYGTRLTRLSITNETFRKTQLYP 269
        |||
Db    241 SSIIDYYGTRLTRLSITNETFRKTQLYP 269

```

# RESULT 2

US-09-833-503A-2

```

; Sequence 2, Application US/09833503A
; Patent No. US20020146760A1
; GENERAL INFORMATION:
; APPLICANT: Ozenberger, Bradley A
; APPLICANT: Kajkowski, Eileen M
; APPLICANT: Lo, Ching-Hsiung F

```

```
; APPLICANT: American Home Products Corporation
; TITLE OF INVENTION: No. US20020146760A1e1 G-Protein-Coupled Receptor-Like
Proteins and
; TITLE OF INVENTION: Polynucleotides Encoded By Them, and Methods of Using
; TITLE OF INVENTION: Same
; FILE REFERENCE: AHP98165-00PCT
; CURRENT APPLICATION NUMBER: US/09/833,503A
; CURRENT FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/104,104
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-503A-2
```

```
Query Match          100.0%; Score 1439; DB 9; Length 269;
Best Local Similarity 100.0%; Pred. No. 3e-135;
Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 MHILKGSPNVIPRAHGQKNTRRDGTGLYPMRGPFKNLALLPFSLP LLGGGGSGSGEKVSV 60
|
Db      1 MHILKGSPNVIPRAHGQKNTRRDGTGLYPMRGPFKNLALLPFSLP LLGGGGSGSGEKVSV 60

Qy     61 SKMAAAWPSGPSAPEAVTARLVGVLFVSVTTGPWGAVATSAGGEESLKCEDLVKGQYIC 120
|
Db     61 SKMAAAWPSGPSAPEAVTARLVGVLFVSVTTGPWGAVATSAGGEESLKCEDLVKGQYIC 120

Qy    121 KDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNG 180
|
Db    121 KDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNG 180

Qy    181 YSYKVAVALSLFLGWLGADRFYLGYPALG LLKFCTVGFCGIGSLIDFILISMQIVGPSDG 240
|
Db    181 YSYKVAVALSLFLGWLGADRFYLGYPALG LLKFCTVGFCGIGSLIDFILISMQIVGPSDG 240

Qy    241 SSYIIDYYGTRLTRLSITNETFRKTQLYP 269
|
Db    241 SSYIIDYYGTRLTRLSITNETFRKTQLYP 269
```

# RESULT 3

US-10-199-881-2

```
; Sequence 2, Application US/10199881
; Publication No. US20030096356A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: No. US20030096356A1e1 G-Protein-Coupled Receptor-Like
Proteins and Polynucleotides
; TITLE OF INVENTION: Encoded by Them, and Methods of Using Same"
; FILE REFERENCE: AHP98165C1
; CURRENT APPLICATION NUMBER: US/10/199,881
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: PCT/ US99/21621
; PRIOR FILING DATE: 1999-10-13
```

```
; PRIOR APPLICATION NUMBER: US 90/833,5081
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: US 60/104,104
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-199-881-2
```

```
Query Match          100.0%; Score 1439; DB 14; Length 269;
Best Local Similarity 100.0%; Pred. No. 3e-135;
Matches 269; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 MHILKGSPNVIPRAHQKNTTRDGTGLYPMRGPFKNLALLPFSLPLLGGGGSGSGEKVSV 60
          |||
Db      1 MHILKGSPNVIPRAHQKNTTRDGTGLYPMRGPFKNLALLPFSLPLLGGGGSGSGEKVSV 60

Qy      61 SKMAAAWPSGPSAPEAVTARLVGVLFVSVTTGPWGAVATSAGGEESLKCEDLKVQYIC 120
          |||
Db      61 SKMAAAWPSGPSAPEAVTARLVGVLFVSVTTGPWGAVATSAGGEESLKCEDLKVQYIC 120

Qy      121 KDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNG 180
          |||
Db      121 KDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNG 180

Qy      181 YSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCTVGFCGIGSLIDFILISMQIVGPSDG 240
          |||
Db      181 YSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCTVGFCGIGSLIDFILISMQIVGPSDG 240

Qy      241 SSYIIDYYGTRLTRLSITNETFRKTQLYP 269.
          |||
Db      241 SSYIIDYYGTRLTRLSITNETFRKTQLYP 269
```

#### RESULT 4

```
US-09-974-879-230
; Sequence 230, Application US/09974879
; Publication No. US20030028003A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 125 Human Secreted Proteins
; FILE REFERENCE: PZ020P2
; CURRENT APPLICATION NUMBER: US/09/974,879
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/239,893
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 09/818,683
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 09/305,736
; PRIOR FILING DATE: 1999-05-05
; PRIOR APPLICATION NUMBER: PCT/US98/23435
; PRIOR FILING DATE: 1998-11-04
; PRIOR APPLICATION NUMBER: US 60/064,911
; PRIOR FILING DATE: 1997-11-07
```

```

; PRIOR APPLICATION NUMBER: US 60/064,912
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,983
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,900
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,988
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,987
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,908
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,984
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/064,985
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: US 60/066,094
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: US 60/066,100
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: US 60/066,089
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: US 60/066,095
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: US 60/066,090
; PRIOR FILING DATE: 1997-11-17
; NUMBER OF SEQ ID NOS: 611
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 230
;   LENGTH: 221
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (184)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-974-879-230

```

```

Query Match          14.0%; Score 201; DB 10; Length 221;
Best Local Similarity 45.7%; Pred. No. 8.4e-12;
Matches 53; Conservative 12; Mismatches 39; Indels 12; Gaps 5;

```

```

Qy      135 CTNYTA--HVSC----FPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAVA 188
      ||| |: ||| :|| | | :| | | | | : | | : |:|
Db      112 CTNSTSCMTVSCPRQRYPA-NCTVRD----HVVHCLGNRT-FPKMLYCJWTGGYKWSTALA 165

Qy      189 LSLFLGWLGADRFYLGYPALGLLKFCTVGFSGIGSLIDFILISMQIVGPSDGSSYI 244
      ||: || ||||| || | : | || :||| :|| : |||: || ||
Db      166 LSITLGGFGADRFYLGQWXEGLGKLFSFGGLGIWTLIDVLLIGVGYVGPADGSLYI 221

```

# RESULT 5

US-09-305-736-230

```

; Sequence 230, Application US/09305736
; Publication No. US20030088078A1
; GENERAL INFORMATION:
; APPLICANT: Feng et al.

```



```

; TITLE OF INVENTION: 125 Human Secreted Proteins
; FILE REFERENCE: PZ020P1
; CURRENT APPLICATION NUMBER: US/09/305,736
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: PCT/US98/23435
; EARLIER FILING DATE: 1998-11-04
; EARLIER APPLICATION NUMBER: 60/064,911
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,912
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,983
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,900
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,988
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,987
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,908
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,984
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/064,985
; EARLIER FILING DATE: 1997-11-07
; EARLIER APPLICATION NUMBER: 60/066,094
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,100
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,089
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,095
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 60/066,090
; EARLIER FILING DATE: 1997-11-17
; NUMBER OF SEQ ID NOS: 612
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 230
; LENGTH: 222
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (184)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (222)
; OTHER INFORMATION: Xaa equals stop translation
US-09-305-736-230

```

```

Query Match          14.0%;  Score 201;  DB 10;  Length 222;
Best Local Similarity 45.7%;  Pred. No. 8.4e-12;
Matches    53;  Conservative    12;  Mismatches    39;  Indels    12;  Gaps    5;

```

```

Qy      135 CTNYTA--HVSC----FPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAVA 188
        ||| |:  ||| :|| || :|  | ||  || :|  || : |:|
Db      112 CTNSTSCMTVSCPRQRYPA-NCTVRD---HVVHCLGNRT-FPKMLYCNWTGGYKWSTALA 165

```

Qy 189 LSLFLGWLGADRFYLGYPALGLLKFC TVGFCGIGSLIDFILISMQIVGPSDGSSYI 244  
 ||: || ||||| || | : | || :|| :|| : |||: || ||  
 Db 166 LSITLGGFGADRFYLGQWXEGLGKLFSFGGLGIWTLIDVLLIGVGYVGPADGSLYI 221

RESULT 6

US-09-818-683-230

; Sequence 230, Application US/09818683  
 ; Publication No. US20030211472A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Feng et al.  
 ; TITLE OF INVENTION: 125 Human Secreted Proteins  
 ; FILE REFERENCE: PZ020P1  
 ; CURRENT APPLICATION NUMBER: US/09/818,683  
 ; CURRENT FILING DATE: 2001-03-28  
 ; Prior application data removed - consult PALM or file wrapper  
 ; NUMBER OF SEQ ID NOS: 612  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 230  
 ; LENGTH: 222  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: SITE  
 ; LOCATION: (184)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; FEATURE:  
 ; NAME/KEY: SITE  
 ; LOCATION: (222)  
 ; OTHER INFORMATION: Xaa equals stop translation  
 US-09-818-683-230

Query Match 14.0%; Score 201; DB 11; Length 222;  
 Best Local Similarity 45.7%; Pred. No. 8.4e-12;  
 Matches 53; Conservative 12; Mismatches 39; Indels 12; Gaps 5;

Qy 135 CTNYTA--HVSC----FPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAVA 188  
 ||| |: ||| :|| | | :| | || | | : | || : |:|  
 Db 112 CTNSTSCMTVSCPRQRYPA-NCTVRD----HVVHCLGNRT-FPKMLYCNWTGGYKWSTALA 165  
 Qy 189 LSLFLGWLGADRFYLGYPALGLLKFC TVGFCGIGSLIDFILISMQIVGPSDGSSYI 244  
 ||: || ||||| || | : | || :|| :|| : |||: || ||  
 Db 166 LSITLGGFGADRFYLGQWXEGLGKLFSFGGLGIWTLIDVLLIGVGYVGPADGSLYI 221

RESULT 7

US-09-833-503A-6

; Sequence 6, Application US/09833503A  
 ; Patent No. US20020146760A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ozenberger, Bradley A  
 ; APPLICANT: Kajkowski, Eileen M  
 ; APPLICANT: Lo, Ching-Hsiung F  
 ; APPLICANT: American Home Products Corporation  
 ; TITLE OF INVENTION: No. US20020146760A1e1 G-Protein-Coupled Receptor-Like  
 Proteins and

```
; TITLE OF INVENTION: Polynucleotides Encoded By Them, and Methods of Using
; TITLE OF INVENTION: Same
; FILE REFERENCE: AHP98165-00PCT
; CURRENT APPLICATION NUMBER: US/09/833,503A
; CURRENT FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/104,104
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-503A-6
```

```
Query Match          13.9%; Score 200; DB 9; Length 221;
Best Local Similarity 45.7%; Pred. No. 1.1e-11;
Matches 53; Conservative 12; Mismatches 39; Indels 12; Gaps 5;
```

```
Qy      135 CTNYTA--HVSC----FPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAVA 188
      ||| |: ||| :|| | | :| | | | | : | | : | |
Db      112 CTNSTSCMTVSCPRQRYPA-NCTVRD----HVCCLGNRT-FPKMLYCNWTGGYKWSTALA 165

Qy      189 LSLFLGWLGA DRFYLGYPALGLLK FCTVGF CGIGSLIDFILISMQIVGPSDGSSYI 244
      ||: || ||||| | | | : | || :|| :|| : ||: ||| ||
Db      166 LSITLGGFGADRFYLGQWREGLGKLFSFGGLGIWTLIDVLLIGVGYVGPADGSLYI 221
```

#### RESULT 8

US-09-992-600A-82

```
; Sequence 82, Application US/09992600A
; Publication No. US20030027161A1
; GENERAL INFORMATION:
; APPLICANT: Benjanin, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US4.DIV
; CURRENT APPLICATION NUMBER: US/09/992,600A
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: JPatent
; SEQ ID NO 82
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Homo sapiens
```

; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: 1..32  
US-09-992-600A-82

Query Match 13.9%; Score 200; DB 10; Length 221;  
Best Local Similarity 45.7%; Pred. No. 1.1e-11;  
Matches 53; Conservative 12; Mismatches 39; Indels 12; Gaps 5;

```
Qy      135 CTNYTA--HVSC----FPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAVA 188
          ||| |: ||| :|| || :| | || | | : | || : |:|
Db      112 CTNSTSCMTVSCPRQRYPA-NCTVRD----HVFCLGNRT-FPKMLYCNWTGGYKWSTALA 165

Qy      189 LSLFLGWLGADRFYLGYPALGLLKFCCTVGFCGIGSLIDFILISMQIVGPSDGSSYI 244
          ||: || ||||| || | : | || :|| :|| : |||: ||| ||
Db      166 LSITLGGFGADRFYLGQWREGLGKLFSFGGLGIWTLIDVLLIGVGYVGPADGSLYI 221
```

RESULT 9

US-09-924-340-82

; Sequence 82, Application US/09924340  
; Publication No. US20030027248A1  
; GENERAL INFORMATION:  
; APPLICANT: Bejanin, Stephane  
; APPLICANT: Tanaka, Hiroaki  
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF  
; FILE REFERENCE: 91.US2.REG  
; CURRENT APPLICATION NUMBER: US/09/924,340  
; CURRENT FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: US 60/305,456  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/302,277  
; PRIOR FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/298,698  
; PRIOR FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: US 60/293,574  
; PRIOR FILING DATE: 2001-05-25  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: JPatent  
; SEQ ID NO 82  
; LENGTH: 221  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: 1..32  
US-09-924-340-82

Query Match 13.9%; Score 200; DB 10; Length 221;  
Best Local Similarity 45.7%; Pred. No. 1.1e-11;  
Matches 53; Conservative 12; Mismatches 39; Indels 12; Gaps 5;

```
Qy      135 CTNYTA--HVSC----FPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAVA 188
          ||| |: ||| :|| || :| | || | | : | || : |:|
Db      112 CTNSTSCMTVSCPRQRYPA-NCTVRD----HVFCLGNRT-FPKMLYCNWTGGYKWSTALA 165

Qy      189 LSLFLGWLGADRFYLGYPALGLLKFCCTVGFCGIGSLIDFILISMQIVGPSDGSSYI 244
```

Db                    ||: ||   |||||   || | : |   || :|| :|| :   ||: || ||  
166 LSITLGGFGADRFYLGQWREGLGKLFSFGGLGIWTLIDVLLIGVGYVGPADGSLYI 221

RESULT 10

US-09-992-095B-82  
; Sequence 82, Application US/09992095B  
; Publication No. US20030157485A1  
; GENERAL INFORMATION:  
; APPLICANT: Benjanin, Stephane  
; APPLICANT: Tanaka, Hiroaki  
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF  
; FILE REFERENCE: 91.US5.DIV  
; CURRENT APPLICATION NUMBER: US/09/992,095B  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: US 09/924,340  
; PRIOR FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: PCT/IB01/01715  
; PRIOR FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: US 60/305,456  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/302,277  
; PRIOR FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/298,698  
; PRIOR FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: US 60/293,574  
; PRIOR FILING DATE: 2001-05-25  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: JPatent  
; SEQ ID NO 82  
; LENGTH: 221  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: 1..32  
US-09-992-095B-82

Query Match                    13.9%;   Score 200;   DB 10;   Length 221;  
Best Local Similarity       45.7%;   Pred. No. 1.1e-11;  
Matches       53;   Conservative       12;   Mismatches       39;   Indels       12;   Gaps       5;

Qy            135 CTNYTA--HVSC----FPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAVA 188  
             ||| |:   |||       :|| | | :|       |   ||   | | : |       || :   |:|  
Db            112 CTNSTSCMTVSCPRQRYPA-NCTVRD----HVVHCLGNRT-FPKMLYCNWTGGYKWSTALA 165  
  
Qy            189 LSLFLGWLGADRFYLGYPALGLLKFCCTVGFCEGIGSLIDFILISMQIVGPSDGSSYI 244  
             ||: ||   |||||       || | : |   || :||| :|| :   ||: || ||  
Db            166 LSITLGGFGADRFYLGQWREGLGKLFSFGGLGIWTLIDVLLIGVGYVGPADGSLYI 221

RESULT 11

US-09-999-570-82  
; Sequence 82, Application US/09999570  
; Publication No. US20030170628A1  
; GENERAL INFORMATION:  
; APPLICANT: Benjanin, Stephane

```

; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: G-091US08DIV
; CURRENT APPLICATION NUMBER: US/09/999,570
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: JPatent
; SEQ ID NO 82
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: 1..32
US-09-999-570-82

```

```

Query Match          13.9%; Score 200; DB 10; Length 221;
Best Local Similarity 45.7%; Pred. No. 1.1e-11;
Matches 53; Conservative 12; Mismatches 39; Indels 12; Gaps 5;

```

```

Qy      135 CTNYTA--HVSC----FPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAVA 188
      ||| |: ||| :|| | | :| | | | | : | | : | :|
Db      112 CTNSTSCMTVSCPRQRYPA-NCTVRD---HVVHCLGNRT-FPKMLYCNWTGGYKWSTALA 165

Qy      189 LSLFLGWLGADRFYLGYPALGLLKFCTVGFCGIGSLIDFILISMQIVGPSDGSSYI 244
      ||: || ||||| || | : | || :|| :|| : |||: || ||
Db      166 LSITLGGFGADRFYLGQWREGLGKLFSFGGLGIWTLIDVLLIGVGYVGPADGSLYI 221

```

# RESULT 12

US-10-000-489-82

```

; Sequence 82, Application US/10000489
; Publication No. US20030092011A1
; GENERAL INFORMATION:
; APPLICANT: Benjanin, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US6.DIV
; CURRENT APPLICATION NUMBER: US/10/000,489
; CURRENT FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456

```

; PRIOR FILING DATE: 2001-07-13  
 ; PRIOR APPLICATION NUMBER: US 60/302,277  
 ; PRIOR FILING DATE: 2001-06-29  
 ; PRIOR APPLICATION NUMBER: US 60/298,698  
 ; PRIOR FILING DATE: 2001-06-15  
 ; PRIOR APPLICATION NUMBER: US 60/293,574  
 ; PRIOR FILING DATE: 2001-05-25  
 ; NUMBER OF SEQ ID NOS: 112  
 ; SOFTWARE: JPatent  
 ; SEQ ID NO 82  
 ; LENGTH: 221  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: SIGNAL  
 ; LOCATION: 1..32  
 US-10-000-489-82

Query Match 13.9%; Score 200; DB 14; Length 221;  
 Best Local Similarity 45.7%; Pred. No. 1.1e-11;  
 Matches 53; Conservative 12; Mismatches 39; Indels 12; Gaps 5;

Qy 135 CTNYTA--HVSC----FPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAVA 188  
 ||| |: ||| :|| || :| | || | | : | || : |:|  
 Db 112 CTNSTSCMTVSCPRQRYPA-NCTVRD----HVCCLGNRT-FPKMLYCNWTGGYKWSTALA 165  
 Qy 189 LSLFLGWLGADRFYLGYPALGLLKFCTVGF CGIGSLIDFILISMQIVGPSDGSSYI 244  
 ||: || ||||| || | : | || :|| :|| : |||: || ||  
 Db 166 LSITLGGFGADRFYLGQWREGLGKLFSFGGLGIWTLIDVLLIGVGYVGPADGSLYI 221

# RESULT 13

US-10-000-986-82

; Sequence 82, Application US/10000986  
 ; Publication No. US20030096247A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Benjanin, Stephane  
 ; APPLICANT: Tanaka, Hiroaki  
 ; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF  
 ; FILE REFERENCE: 91.US9.DIV  
 ; CURRENT APPLICATION NUMBER: US/10/000,986  
 ; CURRENT FILING DATE: 2001-11-14  
 ; PRIOR APPLICATION NUMBER: US 09/924,340  
 ; PRIOR FILING DATE: 2001-08-06  
 ; PRIOR APPLICATION NUMBER: PCT/IB01/01715  
 ; PRIOR FILING DATE: 2001-08-06  
 ; PRIOR APPLICATION NUMBER: US 60/305,456  
 ; PRIOR FILING DATE: 2001-07-13  
 ; PRIOR APPLICATION NUMBER: US 60/302,277  
 ; PRIOR FILING DATE: 2001-06-29  
 ; PRIOR APPLICATION NUMBER: US 60/298,698  
 ; PRIOR FILING DATE: 2001-06-15  
 ; PRIOR APPLICATION NUMBER: US 60/293,574  
 ; PRIOR FILING DATE: 2001-05-25  
 ; NUMBER OF SEQ ID NOS: 112  
 ; SOFTWARE: JPatent  
 ; SEQ ID NO 82

; LENGTH: 221  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: 1..32  
US-10-000-986-82

Query Match 13.9%; Score 200; DB 14; Length 221;  
Best Local Similarity 45.7%; Pred. No. 1.1e-11;  
Matches 53; Conservative 12; Mismatches 39; Indels 12; Gaps 5;

Qy 135 CTNYTA--HVSC----FPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAVA 188  
||| |: ||| :|| || :| | || | | : | || : |:|  
Db 112 CTNSTSCMTVSCPRQRYPA-NCTVRD----HVVHCLGNRT-FPKMLYCNWTGGYKWSTALA 165  
  
Qy 189 LSLFLGWLGADRFYLGYPALGLLKFCCTVGFCGIGSLIDFILISMQIVGPSDGSSYI 244  
||: || ||||| || | : | || :|| :|| : ||: ||| ||  
Db 166 LSITLGGFGADRFYLGQWREGLGKLFSFGGLGIWTLIDVLLIGVGYVGPADGSLYI 221

RESULT 14

US-10-199-881-6

; Sequence 6, Application US/10199881  
; Publication No. US20030096356A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; TITLE OF INVENTION: No. US20030096356A1e1 G-Protein-Coupled Receptor-Like  
Proteins and Polynucleotides  
; TITLE OF INVENTION: Encoded by Them, and Methods of Using Same"  
; FILE REFERENCE: AHP98165C1  
; CURRENT APPLICATION NUMBER: US/10/199,881  
; CURRENT FILING DATE: 2002-07-18  
; PRIOR APPLICATION NUMBER: PCT/ US99/21621  
; PRIOR FILING DATE: 1999-10-13  
; PRIOR APPLICATION NUMBER: US 90/833,5081  
; PRIOR FILING DATE: 2001-12-04  
; PRIOR APPLICATION NUMBER: US 60/104,104  
; PRIOR FILING DATE: 1998-10-13  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6  
; LENGTH: 221  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-199-881-6

Query Match 13.9%; Score 200; DB 14; Length 221;  
Best Local Similarity 45.7%; Pred. No. 1.1e-11;  
Matches 53; Conservative 12; Mismatches 39; Indels 12; Gaps 5;

Qy 135 CTNYTA--HVSC----FPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAVA 188  
||| |: ||| :|| || :| | || | | : | || : |:|  
Db 112 CTNSTSCMTVSCPRQRYPA-NCTVRD----HVVHCLGNRT-FPKMLYCNWTGGYKWSTALA 165  
  
Qy 189 LSLFLGWLGADRFYLGYPALGLLKFCCTVGFCGIGSLIDFILISMQIVGPSDGSSYI 244  
||: || ||||| || | : | || :|| :|| : ||: ||| ||



Db 166 LSITLGGFGADRFYLGQWREGLGKLFSFGGLGIWTLIDVLLIGVGYVGPADGSLYI 221

RESULT 15

US-10-154-678-82

; Sequence 82, Application US/10154678  
; Publication No. US20030162186A1  
; GENERAL INFORMATION:  
; APPLICANT: Benjanin, Stephane  
; APPLICANT: Tanaka, Hiroaki  
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF  
; FILE REFERENCE: 182.US1.REG  
; CURRENT APPLICATION NUMBER: US/10/154,678  
; CURRENT FILING DATE: 2002-10-15  
; PRIOR APPLICATION NUMBER: US 09/924,340  
; PRIOR FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: US 60/305,456  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/302,277  
; PRIOR FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/298,698  
; PRIOR FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: US 60/293,574  
; PRIOR FILING DATE: 2001-05-25  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: JPatent  
; SEQ ID NO 82  
; LENGTH: 221  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: -32..-1  
US-10-154-678-82

Query Match 13.9%; Score 200; DB 14; Length 221;  
Best Local Similarity 45.7%; Pred. No. 1.1e-11;  
Matches 53; Conservative 12; Mismatches 39; Indels 12; Gaps 5;

QY 135 CTNYTA--HVSC---FPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAVA 188  
||| |: ||| :|| || :| | || | | : | || : |:|  
Db 112 CTNSTSCMTVSCPRQRYPA-NCTVRD---HVCCLGNRT-FPKMLYCNWTGGYKWSTALA 165  
QY 189 LSLFLGWLGADRFYLGYPALGLLKFCTVGFCGIGSLIDFILISMQIVGPSDGSSYI 244  
||: || ||||| || | : | || :|| :|| : |||: || ||  
Db 166 LSITLGGFGADRFYLGQWREGLGKLFSFGGLGIWTLIDVLLIGVGYVGPADGSLYI 221

Search completed: March 4, 2004, 10:46:22  
Job time : 653 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2004, 10:13:25 ; Search time 80 Seconds  
(without alignments)  
1060.930 Million cell updates/sec

Title: US-09-852-100B-2  
Perfect score: 1439  
Sequence: 1 MHILKGSPNVIPRAHGQKNT.....TRLTRLSITNETFRKTQLYP 269

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertibrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	Match	Length	DB	ID	Description
No.							

---

1	1113	77.3	207	4	Q9BX74	Q9bx74 homo sapien
2	941.5	65.4	208	11	Q99MB3	Q99mb3 mus musculu
3	338	23.5	178	5	Q9W2H1	Q9w2h1 drosophila
4	278	19.3	329	5	Q95PJ8	Q95pj8 caenorhabdi
5	200	13.9	221	4	Q9H651	Q9h651 homo sapien
6	200	13.9	247	4	Q9BRN9	Q9brn9 homo sapien
7	198.5	13.8	261	11	Q8BJ83	Q8bj83 mus musculu
8	196	13.6	230	11	Q9D156	Q9d156 mus musculu
9	182	12.6	284	5	Q9U4H5	Q9u4h5 drosophila
10	172	12.0	80	4	Q9H046	Q9h046 homo sapien
11	169.5	11.8	195	5	Q95QZ5	Q95qz5 caenorhabdi
12	136.5	9.5	214	4	Q9BX73	Q9bx73 homo sapien
13	125	8.7	172	5	Q9VY86	Q9vy86 drosophila
14	121	8.4	149	4	Q9BSR6	Q9bsr6 homo sapien
15	121	8.4	171	4	Q8N0X9	Q8n0x9 homo sapien
16	117	8.1	106	2	Q9S022	Q9s022 borrelia bu
17	116.5	8.1	213	11	Q9CWL9	Q9cwl9 mus musculu
18	116.5	8.1	213	11	Q8R0I4	Q8r0i4 mus musculu
19	116.5	8.1	213	11	Q8BJJ1	Q8bjj1 mus musculu
20	111	7.7	184	16	Q87V96	Q87v96 pseudomonas
21	107	7.4	106	16	Q8F808	Q8f808 leptospira
22	105.5	7.3	142	16	Q8DJB9	Q8djb9 synechococc
23	105	7.3	89	2	Q9S015	Q9s015 borrelia bu
24	105	7.3	132	9	Q854V8	Q854v8 mycobacteri
25	104	7.2	159	16	Q8FU99	Q8fu99 corynebacte
26	98	6.8	1092	5	Q964R2	Q964r2 theileria t
27	97	6.7	137	16	Q932P1	Q932p1 staphylococ
28	96.5	6.7	765	16	Q7V2C8	Q7v2c8 prochloroco
29	95.5	6.6	913	4	O75061	O75061 homo sapien
30	95	6.6	139	16	Q88CT9	Q88ct9 pseudomonas
31	95	6.6	309	16	Q9RS03	Q9rs03 deinococcus
32	94.5	6.6	166	16	Q82HV7	Q82hv7 streptomyce
33	92.5	6.4	76	16	Q8EVR7	Q8evr7 mycoplasma
34	92.5	6.4	87	16	Q82HV9	Q82hv9 streptomyce
35	92	6.4	456	12	Q66668	Q66668 equine herp
36	92	6.4	720	5	Q9V5W4	Q9v5w4 drosophila
37	92	6.4	720	5	Q961T2	Q961t2 drosophila
38	91.5	6.4	168	16	Q7URJ9	Q7urj9 rhodopirell
39	91	6.3	1324	12	Q9J3E7	Q9j3e7 murine hepa
40	90.5	6.3	151	11	Q8K181	Q8k181 mus musculu
41	90	6.3	204	10	Q8RZE7	Q8rze7 oryza sativ
42	90	6.3	443	16	Q82M28	Q82m28 streptomyce
43	89.5	6.2	690	10	O22512	O22512 glycine max
44	89.5	6.2	792	16	Q7VD31	Q7vd31 prochloroco
45	89	6.2	330	12	Q98339	Q98339 murine hepa

#### ALIGNMENTS

##### RESULT 1

Q9BX74

ID Q9BX74 PRELIMINARY; PRT; 207 AA.  
AC Q9BX74;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Beta-amyloid binding protein.  
 GN BBP.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21276355; PubMed=11278849;  
 RA Kajkowski E.M., Lo C.F., Ning X., Walker S., Sofia H.J., Wang W.,  
 RA Edris W., Chanda P., Wagner E., Vile S., Ryan K., McHendry-Rinde B.,  
 RA Smith S.C., Wood A., Rhodes K.J., Kennedy J.D., Bard J.,  
 RA Jacobsen J.S., Ozenberger B.A.;  
 RT "beta-Amyloid Peptide-induced Apoptosis Regulated by a Novel Protein  
 RT Containing a G Protein Activation Module.";  
 RL J. Biol. Chem. 276:18748-18756(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RA Strausberg R.;  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF353990; AAK35064.1; -.  
 DR EMBL; BC029486; AAH29486.1; -.  
 DR InterPro; IPR007829; TM2.  
 DR Pfam; PF05154; TM2; 1.  
 KW Signal.  
 FT SIGNAL 1 37 POTENTIAL.  
 SQ SEQUENCE 207 AA; 22326 MW; A5590FD7AECDF292 CRC64;

Query Match 77.3%; Score 1113; DB 4; Length 207;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-94;  
 Matches 207; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      63 MAAAWPSGPSAPEAVTARLVGVLFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYICKD 122
      |||
Db      1 MAAAWPSGPSAPEAVTARLVGVLFVSVTTGPWGAVATSAGGEESLKCEDLKVGQYICKD 60

QY     123 PKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYS 182
      |||
Db      61 PKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYS 120

QY     183 YKVAVALSLFLGWLGADRFYLGYPALGLLKFCTVGFCGIGSLIDFILISMQIVGPSDGSS 242
      |||
Db     121 YKVAVALSLFLGWLGADRFYLGYPALGLLKFCTVGFCGIGSLIDFILISMQIVGPSDGSS 180

QY     243 YIIDYYGTRLTRLSITNETFRKTQLYP 269
      |||
Db     181 YIIDYYGTRLTRLSITNETFRKTQLYP 207

```

# RESULT 2

Q99MB3

ID Q99MB3 PRELIMINARY; PRT; 208 AA.  
 AC Q99MB3;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

```

DE      Beta-amyloid binding protein.
GN      BBP.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=BALB/c;
RX      MEDLINE=21276355; PubMed=11278849;
RA      Kajkowski E.M., Lo C.F., Ning X., Walker S., Sofia H.J., Wang W.,
RA      Edris W., Chanda P., Wagner E., Vile S., Ryan K., McHendry-Rinde B.,
RA      Smith S.C., Wood A., Rhodes K.J., Kennedy J.D., Bard J.,
RA      Jacobsen J.S., Ozenberger B.A.;
RT      "beta-Amyloid Peptide-induced Apoptosis Regulated by a Novel Protein
RT      Containing a G Protein Activation Module.";
RL      J. Biol. Chem. 276:18748-18756(2001).
DR      EMBL; AF353993; AAK35067.1; -.
DR      MGD; MGI:2137022; Bbp.
DR      GO; GO:0005887; C:integral to plasma membrane; IDA.
DR      GO; GO:0001540; F:beta-amyloid binding; IPI.
DR      GO; GO:0004930; F:G-protein coupled receptor activity; IDA.
DR      GO; GO:0008624; P:induction of apoptosis by extracellular sig. . .; IDA.
DR      InterPro; IPR007829; TM2.
DR      Pfam; PF05154; TM2; 1.
SQ      SEQUENCE      208 AA;  22271 MW;  91A7932163F4F04C CRC64;

```

Qy	63	MAAAWPSGSPSAPEAVTARLVGLWFVSVTTGPWGAVATSA-GGEESLKCEDLKVGYICK	121
		:   : :     :     :       :       :	
Db	1	MAAAWPAGRASPAAGPPGLRLTLWLVTVAAGHCGAAASGAVGGEETPKCEDLRVGQYICK	60
Qy	122	DPKINDATQEVPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGY	181
		:  :	
Db	61	EPKINDATQEVPVNCTNYTAHVQCFCPAPKITCKDLSGNETHFTGSEVGFLKPISCRNVNGY	120
Qy	182	SYKVAVALSLFLGWLGADRFYLGYPALGLLLKFCTVGFCGIGSLIDFILISMQIVGPSDGS	241
Db	121	SYKVAVALSLFLGWLGADRFYLGYPALGLLLKFCTVGFCGIGSLIDFILISMQIVGPSDGS	180
Qy	242	SYIIDYYGTRLTRLSITNETFRKTQLYP	269
Db	181	SYIIDYYGTRLTRLSITNETFRKTQLYP	208

OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; *Drosophila*.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
 RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE003453; AAF46720.1; -.

DR EMBL; AY061343; AAL28891.1; -.  
 DR FlyBase; FBgn0034626; CG10795.  
 DR InterPro; IPR007829; TM2.  
 DR Pfam; PF05154; TM2; 1.  
 SQ SEQUENCE 178 AA; 19896 MW; 17C41166607ACC03 CRC64;

Query Match 23.5%; Score 338; DB 5; Length 178;  
 Best Local Similarity 42.6%; Pred. No. 3.8e-23;  
 Matches 69; Conservative 30; Mismatches 49; Indels 14; Gaps 5;

Qy 107 SLKCEDLK-VGQYICKDP---KINDATQEPVNCTNY-TAHVSCFPAPNITCKDSSGNETH 161  
 :: | :|: :||:| || :|: ||: || | | | | : :|| |  
 Db 20 NVDCNELQMMGQFMCPDPARGQIDPKTQQLAGCTREGRARVWCIAANEINCTE-TGNAT- 77

Qy 162 FTGNEVGFFKPISCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCTVGFCGI 221  
 | : : | : || : ||: || | ||||| : ||||| : | :  
 Db 78 -----FTREVPCKWTNGYHLDITLLLSVFLGMFGVDRFYLGYPGIGLLKFCTLGGMFL 130

Qy 222 GSLIDFILISMQIVGPSDGSSYIIDYYGTRLRLSITNETFR 263  
 | ||| :||:|:| ||:| ||| : : | |:|  
 Db 131 GQLIDIVLIALQVVGPADGSAYVIPYYGAGIHIVRSNDNTTYR 172

#### RESULT 4

Q95PJ8

ID Q95PJ8 PRELIMINARY; PRT; 329 AA.  
 AC Q95PJ8;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Y66D12A.21 protein.  
 GN Y66D12A.21.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Sulston J.E.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 RT investigating biology.";  
 RL Science 282:2012-2018(1998).  
 DR EMBL; AL161712; CAC35892.1; -.  
 DR WormPep; Y66D12A.21; CE26465.  
 DR InterPro; IPR007829; TM2.  
 DR Pfam; PF05154; TM2; 1.  
 SQ SEQUENCE 329 AA; 38167 MW; 9C6FB3EE7E3866D0 CRC64;

Query Match 19.3%; Score 278; DB 5; Length 329;  
 Best Local Similarity 36.6%; Pred. No. 2.8e-17;  
 Matches 63; Conservative 34; Mismatches 59; Indels 16; Gaps 5;

Qy 98 VATSAGGEESLKCEDLKVGQYICKDPKINDATQEPVNC-TNYTAHVSCFPAPNITC--KD 154  
 :: || : ::||:| | |::|: :: | | : : | | : | | ||  
 Db 11 ISVSA-SDATVKCDDLDPNQYLCKNYAVDTITQQSVTCAADNSIQVMCETAEHKCVGKD 69  
 Qy 155 SSG--NETHFTGNEVGFFKPISCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALGLLK 212  
 | | | | : | | | | | : | | | | | : |  
 Db 70 QFGIFNRT-----VPSACHYGAHVSYTTTVLLSIFLGFFGIDRIYLYGYALGLIK 119  
 Qy 213 FCTVGFCGIGSLIDFILISMQIVGPSDGSSYIIDYYGTRLRLSITNETFRK 264  
 ::| : | | | | : | | : | | : : : : |  
 Db 120 MFSLGGLFVFWLVLDIILISLQLLGPADGTAYAMAYYGPKAQMIRLVATIWK 171

RESULT 5

Q9H651

ID Q9H651 PRELIMINARY; PRT; 221 AA.  
 AC Q9H651;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein FLJ22604 (BBP-like protein 2).  
 GN BLP2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Small intestine;  
 RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,  
 RA Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,  
 RA Nakamura Y., Isogai T., Sugano S.;  
 RT "NEDO human cDNA sequencing project.";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Kajkowski E.M., Lo C.F., Ning X., Walker S., Sofia H.J., Wang W.,  
 RA Edris W., Chanda P., Wagner E., Vile S., Ryan K., McHendry-Rinde B.,  
 RA Smith S.C., Wood A., Rhodes K.J., Kennedy J.D., Bard J.,  
 RA Jacobsen J.S., Ozenberger B.A.;  
 RT "Beta-amyloid peptide-induced apoptosis regulated by a novel protein  
 RT containing a G protein activation module.";  
 RL J. Biol. Chem. 0:0-0(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Muscle;  
 RA Strausberg R.;  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AK026257; BAB15415.1; -.  
 DR EMBL; AF353992; AAK35066.1; -.  
 DR EMBL; BC008873; AAH08873.1; -.  
 DR InterPro; IPR007829; TM2.  
 DR Pfam; PF05154; TM2; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 221 AA; 24410 MW; 92151D6EF6363D74 CRC64;

Query Match

13.9%; Score 200; DB 4; Length 221;



Best Local Similarity 45.7%; Pred. No. 2.6e-10;  
Matches 53; Conservative 12; Mismatches 39; Indels 12; Gaps 5;

```
Qy      135 CTNYTA--HVSC----FPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAVA 188
      ||| |: ||| :|| || :| | || | | : | || : |:|
Db      112 CTNSTSCMTVSCPRQRYPA-NCTVRD----HVVHCLGNRT-FPKMLYCNWTGGYKWSTALA 165

Qy      189 LSLFLGWLGADRFYLGYPALGGLLKFCVTGFCGIGSLIDFILISMQIVGPSDGSSSYI 244
      ||: || ||||| || | : | || :|| :|| : |||: || ||
Db      166 LSITLGGFGADRFYLGQWREGLGKLFSFGGLGIWTLIDVLLIGVGYVGPADGSLYI 221
```

RESULT 6

Q9BRN9

ID Q9BRN9 PRELIMINARY; PRT; 247 AA.  
AC Q9BRN9;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Similar to hypothetical protein FLJ22604.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Colon;  
RA Strausberg R.;  
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; BC006150; AAH06150.1; -.  
DR InterPro; IPR007829; TM2.  
DR Pfam; PF05154; TM2; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 247 AA; 27161 MW; CE1D0D9C53DDF73C CRC64;

Query Match 13.9%; Score 200; DB 4; Length 247;  
Best Local Similarity 45.7%; Pred. No. 3e-10;  
Matches 53; Conservative 12; Mismatches 39; Indels 12; Gaps 5;

```
Qy      135 CTNYTA--HVSC----FPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYKVAVA 188
      ||| |: ||| :|| || :| | || | | : | || : |:|
Db      138 CTNSTSCMTVSCPRQRYPA-NCTVRD----HVVHCLGNRT-FPKMLYCNWTGGYKWSTALA 191

Qy      189 LSLFLGWLGADRFYLGYPALGGLLKFCVTGFCGIGSLIDFILISMQIVGPSDGSSSYI 244
      ||: || ||||| || | : | || :|| :|| : |||: || ||
Db      192 LSITLGGFGADRFYLGQWREGLGKLFSFGGLGIWTLIDVLLIGVGYVGPADGSLYI 247
```

RESULT 7

Q8BJ83

ID Q8BJ83 PRELIMINARY; PRT; 261 AA.  
AC Q8BJ83;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Similar to BBP-like protein 2.  
GN 5930422O05RIK.

OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Forelimb;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK077858; BAC37037.1; -.  
 DR MGD; MGI:1924429; 5930422O05Rik.  
 DR InterPro; IPR007829; TM2.  
 DR Pfam; PF05154; TM2; 1.  
 SQ SEQUENCE 261 AA; 28880 MW; 70346780D3CF5CDB CRC64;

Query Match 13.8%; Score 198.5; DB 11; Length 261;  
 Best Local Similarity 26.8%; Pred. No. 4.4e-10;  
 Matches 69; Conservative 23; Mismatches 86; Indels 79; Gaps 6;

```

Qy      46 LLGGGGSGSGEKVSVSKMAAAWPSGPSAPEAVTARLVGVLFVSVTTGPWGAVATSAGGE 105
      :| | | | : |      | | | : | |      | :|
Db      26 ILSGDGSLNLEHSQPLAQAIKDP-GPTRTFSVVPRAAENQLFSHLT----- 70

Qy      106 ESLKCEDLKVGGYICKDPKINDATQEPVNCTNYTAHVS-----CFPAPNITCKDS 155
      | : : | : | |      : : | : |      : ||      | | : : | |
Db      71 -----ESTEIPPYMTKCPSNGLCSRLPADCIECATNVSCTYGKPVTFDCTVKPSVTCVDQ 125

Qy      156 S-----GNET-----HFTGNEV 167
      | |      | |
Db      126 DLKPQRNFVINMTCRFCWQLPETDYECSNSTTCMTVACPRQRYFANCTVRDHIHCLGNRT 185

Qy      168 GFFKPISCRNVNGYSYKVAVALSLFLGWLGADR FYLGYPALGLLKFCTVGFCGIGSLIDF 227
      | | : |      | | : | : | : | | | | | | | | | | | : | | : | |
Db      186 -FPKLLYCNWTGGYKWSTRALALSITLGGFGADR FYLGQWREGLGKLFSGGLGIWTLIDV 244

Qy      228 ILISMQIVGPSDGSSYI 244
      : | | : | | : | | | |
Db      245 LLIGVGYVGPADGSLYI 261

```

# RESULT 8

Q9D156

ID Q9D156 PRELIMINARY; PRT; 230 AA.  
 AC Q9D156;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE 1110025I09Rik protein (RIKEN cDNA 1110025I09 gene).  
 GN BLP2 OR 1110025I09RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Embryo;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Salivary gland;  
 RA Strausberg R.;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AK003917; BAB23075.1; -.  
 DR EMBL; BC024620; AAH24620.1; -.  
 DR MGD; MGI:1915884; Blp2.  
 DR InterPro; IPR007829; TM2.  
 DR Pfam; PF05154; TM2; 1.  
 SQ SEQUENCE 230 AA; 25639 MW; 396D650D8BEE99A5 CRC64;

## RESULT 9

Q9U4H5

ID Q9U4H5 PRELIMINARY; PRT; 284 AA.  
AC Q9U4H5; Q9W361;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE BCDNA.GH02974 (ALMONDEX) (AMX protein).  
GN AMX OR BCDNA:GH02974 OR CG12127.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Rubin G.M., Wan K.H., Harvey D., Lewis S.E., Brokstein P., Tsang G.,  
RA Agbayani A., Arcaina T.T., Baxter E., Blazej R.G., Butenhoff C.,  
RA Champe M., Chavez C., Chew M., Doyle C.M., Farfan D.E., Frise E.,  
RA Galle R., George R.A., Harris N.L., Hoskins R.A., Evans-Holm M.,  
RA Houston K.A., Hummasti S.R., Kim E., Li P., Moshrefi M., Pacleb J.M.,  
RA Park S., Sequeira A., Sethi H., Snir E., Svirskas R.R., Weinburg T.,  
RA Celniker S.E.;  
RT "Full length Drosophila melanogaster cDNA sequence.";  
RL Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Ovary;  
RA Michellod M.-A.E., Remillieux N.C., Randsholt N.B.;  
RT "Characterization of almondex.";  
RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkeley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL; AF181623; AAD55409.1; -.  
 DR EMBL; AF217797; AAF36924.2; -.  
 DR EMBL; AE003446; AAF46474.2; -.  
 DR FlyBase; FBgn0000077; amx.  
 DR GO; GO:0007498; P:mesoderm development; IMP.  
 DR InterPro; IPR001304; Lectin\_C.  
 DR InterPro; IPR007829; TM2.  
 DR Pfam; PF05154; TM2; 1.  
 DR PROSITE; PS00615; C\_TYPE\_LECTIN\_1; 1.  
 SQ SEQUENCE 284 AA; 31364 MW; 8FB8FFB5733AC851 CRC64;

Query Match 12.6%; Score 182; DB 5; Length 284;

Best Local Similarity 33.3%; Pred. No. 1.6e-08;

Matches 50; Conservative 21; Mismatches 51; Indels 28; Gaps 5;

QY 104 GEESLK-----CEDLKVGQYICKDPKINDATQEPVNCTNYTAH--VSCFPAPNITCKD 154  
 ||| : | :: | | :: : : ||| | | |  
 Db 154 GERSFQRMNCRYCYQTEMWQSCGQRSSCNSATDKLFRTNCTVHHDVLCL----- 204  
 QY 155 SSGNETHFTGNEVGFFKPISCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALGLLKFC 214  
 || : || | : | || : | : : || || || || || || : | : |  
 Db 205 --GNRS-FTRN-----LRNWTQGYRWSTALLISLTLGGFGADRFYLGHWQEGIGKLF 254  
 QY 215 TVGFCGIGSLIDFILISMQIVGPSDGSSYI 244  
 : | | : : || : || | : || : || | |  
 Db 255 SFGGLGVWTIIDVLLISMHYLGPADGSLYI 284

# RESULT 10

Q9H046

ID Q9H046 PRELIMINARY; PRT; 80 AA.

AC Q9H046;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Hypothetical protein (Fragment).

GN DKFZP667C1011.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lymph node;  
 RA Koehler K., Beyer A., Mewes H.W., Weil B., Wiemann S.;  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AL512689; CAC21647.1; -.  
 DR InterPro; IPR007829; TM2.  
 DR Pfam; PF05154; TM2; 1.  
 KW Hypothetical protein.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 80 AA; 8699 MW; 8BE6BE788235C58D CRC64;

Query Match 12.0%; Score 172; DB 4; Length 80;  
 Best Local Similarity 46.6%; Pred. No. 2.7e-08;  
 Matches 41; Conservative 9; Mismatches 30; Indels 8; Gaps 1;

QY 157 GNETHFTGNEVGFFKPISCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCTV 216  
 || | : | : : || : || || || || || || || :  
 Db 1 GNRT-----FPKMLYCNWTGGYKWSTALALSITLGGFGADRFYLGQWREGLGKLFSF 52  
 QY 217 GFCGIGSLIDFILISMQIVGPSDGSSYI 244  
 | || : || : || : || : || || || ||  
 Db 53 GGLGIWTLIDVLLIGVGYVGPADGSLYI 80

# RESULT 11

Q95QZ5

ID Q95QZ5 PRELIMINARY; PRT; 195 AA.

AC Q95QZ5;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein.

GN C41D11.9.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RX MEDLINE=99069613; PubMed=9851916;

RA None;

RT "Genome sequence of the nematode C. elegans: a platform for

RT investigating biology. The C. elegans Sequencing Consortium.";

RL Science 282:2012-2018(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Gattung S., Maggi L.;

RT "The sequence of C. elegans cosmid C41D11.";

RL Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Waterston R.;

RT "Direct Submission.";

RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF003740; AAL08031.1; -.  
 DR WormPep; C41D11.9; CE29489.  
 DR InterPro; IPR007829; TM2.  
 DR Pfam; PF05154; TM2; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 195 AA; 21203 MW; 35945E407F184DAE CRC64;

Query Match 11.8%; Score 169.5; DB 5; Length 195;  
 Best Local Similarity 28.9%; Pred. No. 1.4e-07;  
 Matches 48; Conservative 27; Mismatches 62; Indels 29; Gaps 5;

Qy 104 GEESLKCE---DLKVGQYICKDPKINDATQE----PVNCTNYTA-----HVSCF 145  
 | | | | : : | : | : : | | : | | |  
 Db 34 GSAGLTCTFPGD CRIGDTV----KVNCTSRKGCPNPVSRNNVEAVCRFCWQLLP GDYDCE 89  
 Qy 146 PAPNITCKDS-----SGNETHFTGNEVGFFKPISCRNVNGYSYKVAVALSLFLGWLGA 198  
 || | : : | : : : | : | | : || : : || : || ||  
 Db 90 PATNCSTSSTKLLVTKCSAHSSVICMGQRNFYKRIPCNWSSGYSWTKTMILSVVLGGFGA 149  
 Qy 199 DRFYLGYPALGLLKFCTVGFCCGIGSLIDFILISMQIVGPSDGSSYI 244  
 ||||| : | : | | : : | : | : : : | || ||  
 Db 150 DRFYLGWLKSAIGKLFSFGGLGVWTLVDVVLIAVGYIKPYDGS MYI 195

# RESULT 12

Q9BX73

ID Q9BX73 PRELIMINARY; PRT; 214 AA.  
 AC Q9BX73;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE BBP-like protein 1.  
 GN BLP1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21276355; PubMed=11278849;  
 RA Kajkowski E.M., Lo C.F., Ning X., Walker S., Sofia H.J., Wang W.,  
 RA Edris W., Chanda P., Wagner E., Vile S., Ryan K., McHendry-Rinde B.,  
 RA Smith S.C., Wood A., Rhodes K.J., Kennedy J.D., Bard J.,  
 RA Jacobsen J.S., Ozenberger B.A.;  
 RT "beta-Amyloid Peptide-induced Apoptosis Regulated by a Novel Protein  
 RT Containing a G Protein Activation Module."  
 RL J. Biol. Chem. 276:18748-18756(2001).  
 DR EMBL; AF353991; AAK35065.1; -.  
 DR InterPro; IPR007829; TM2.  
 DR Pfam; PF05154; TM2; 1.  
 SQ SEQUENCE 214 AA; 22871 MW; BB928712AF2F78A8 CRC64;

Query Match 9.5%; Score 136.5; DB 4; Length 214;  
 Best Local Similarity 27.8%; Pred. No. 0.00018;  
 Matches 58; Conservative 21; Mismatches 83; Indels 47; Gaps 10;

Qy 59 SVSKMAAAMP----SGPSAPEAVTARLVGVLFVSVTTGPWGAVATSAGGEES--LKCED 112  
 | : | | | : | : | | | | : | | : |  
 Db 33 SHSQNATAEPELTSAGAAQPE-----GPGGAASWEYGDPHSPVILCSY 75

Qy 113 LKVGQYICKDP--KINDAT--QE-PVNCTNYTAH-----VSCFPAPNITCKDSSGN 158  
 | : | | : : | | | : | | | | :  
 Db 76 LPDEFIECEDPVDHVGNATASQELGYGCLKFGGQAYS DVEHTSVQCHALDGI EC---ASP 132

Qy 159 ETHFTGNEVGFFKPISCRNVNGYSYKVAVALSLFLGWL GADRFYLGYPALGLLK FCTVGF 218  
 | | | | | : : : | | | | | : | : |  
 Db 133 RTFLREN-----KP--CIKYTGHYFITLLYSFFLGCGFVDRFCLGHTGTAVGKLLTLGG 185

Qy 219 CGIGSLIDFILISMQIVGPSDGS SYIIDY 247  
 | | : | | : : | | | : |  
 Db 186 LGIWWFVDLILLITGGLMPSDGSNWCTVY 214

# RESULT 13

Q9VY86

ID Q9VY86 PRELIMINARY; PRT; 172 AA.  
 AC Q9VY86;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE CG11103 protein (LP03404p).  
 GN CG11103.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,



RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of *Drosophila melanogaster* genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of *Drosophila melanogaster* genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,

RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,  
 RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,  
 RA Patel S., Phouanenvong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
 RA Celniker S.;  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AE003493; AAF48318.2; -.  
 DR EMBL; AY119007; AAM50867.1; -.  
 DR FlyBase; FBgn0030522; CG11103.  
 DR InterPro; IPR007829; TM2.  
 DR Pfam; PF05154; TM2; 1.  
 SQ SEQUENCE 172 AA; 18809 MW; 73DFFB4B1B780E39 CRC64;

Query Match 8.7%; Score 125; DB 5; Length 172;  
 Best Local Similarity 25.9%; Pred. No. 0.0015;  
 Matches 38; Conservative 18; Mismatches 49; Indels 42; Gaps 5;

Qy 120 CKDP-----KINDATQEPVN-----CTNYTAHVSCFPAPNITCKDSS 156  
 |||| | :| | | | : | : | :  
 Db 41 CKDPVDHRENATAQQEKYKGLKFGGSTYEEVEHAMVWCTVF-ADIECY----- 88  
  
 Qy 157 GNETHFTGNEVGFFKPISCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALGLLKFC TV 216  
 || | : | : : : | : || : || || | : | :  
 Db 89 GNRTFLRAG-----VPCVRYTDHYFVTTLIYSMLLGLGMDRFLGQTGTAVGKLLTM 141  
  
 Qy 217 GFCGIGSLIDFILISMQIVGPSDGSSY 243  
 | | : || || : | ||| :  
 Db 142 GGVGVWWIIDVILLITNNLLPEDGSNW 168

#### RESULT 14

##### Q9BSR6

ID Q9BSR6 PRELIMINARY; PRT; 149 AA.  
 AC Q9BSR6;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Similar to RIKEN cDNA 2410018G23 gene.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreas;  
 RA Strausberg R.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; BC004878; AAH04878.1; -.  
 DR InterPro; IPR007829; TM2.  
 DR Pfam; PF05154; TM2; 1.  
 SQ SEQUENCE 149 AA; 16490 MW; 749C1C574CB2A518 CRC64;

Query Match 8.4%; Score 121; DB 4; Length 149;  
 Best Local Similarity 29.6%; Pred. No. 0.003;  
 Matches 42; Conservative 16; Mismatches 60; Indels 24; Gaps 7;

Qy 120 CKDP--KINDAT--QE-PVNCTNYTAH-----VSCFPAPNITCKDSSGNETHFTGN 165  
 |:| : :| | | : | | | : | |

```

Db      18 CEDPVDHVG NATASQELGYGCLKFGGQAYS DVEHTSVQCHALD GIEC---ASPRTFLREN 74
Qy      166 EVGFFKPI SCRN VNGYSYKVAVALSLFLGWL GADRFYLGYPALG LLKFCTVGF CGIGSLI 225
      || |   |: :   : | || | ||| ||:   : | |: | ||   :
Db      75 -----KP--CIKYTGHYFIT TLLYSF FLGCFGVDRFCLGHTGTAVGKLLTLGGLGIWWFV 127
Qy      226 DFILISMQIVGPSD GSSYIIDY 247
      | ||:   : |||||::   |
Db      128 DLILLITGGLMPSDG SNWCTVY 149

```

RESULT 15

Q8NOX9

```

ID      Q8NOX9          PRELIMINARY;          PRT;      171 AA.
AC      Q8NOX9;
DT      01-OCT-2002 (TrEMBLrel. 22, Created)
DT      01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT      01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE      Hypothetical protein FLJ90546 (Hypothetical protein FLJ90674).
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Ovarian carcinoma, and Placenta;
RA      Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
RA      Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA      Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
RA      Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA      Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
RT      "NEDO human cDNA sequencing project.";
RL      Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR      EMBL; AK075027; BAC11359.1; -.
DR      EMBL; AK075155; BAC11437.1; -.
DR      InterPro; IPR007829; TM2.
DR      Pfam; PF05154; TM2; 1.
KW      Hypothetical protein.
SQ      SEQUENCE      171 AA;  19012 MW;  CD9F9F7CEE5CB388 CRC64;

```

```

Query Match          8.4%;  Score 121;  DB 4;  Length 171;
Best Local Similarity 29.6%;  Pred. No. 0.0035;
Matches      42;  Conservative 16;  Mismatches 60;  Indels 24;  Gaps 7;

```

```

Qy      120 CKDP--KINDAT--QE-PVNCTNYTAH-----VSCFPAPNITCKDSSGNETHFTGN 165
      |:||   : :|| ||   | :           | |   | |   : |   |
Db      40 CEDPVDHVG NATASQELGYGCLKFGGQAYS DVEHTSVQCHALD GIEC---ASPRTFLREN 96
Qy      166 EVGFFKPI SCRN VNGYSYKVAVALSLFLGWL GADRFYLGYPALG LLKFCTVGF CGIGSLI 225
      || |   |: :   : | || | ||| ||:   : | |: | ||   :
Db      97 -----KP--CIKYTGHYFIT TLLYSF FLGCFGVDRFCLGHTGTAVGKLLTLGGLGIWWFV 149
Qy      226 DFILISMQIVGPSD GSSYIIDY 247
      | ||:   : |||||::   |
Db      150 DLILLITGGLMPSDG SNWCTVY 171
Search completed: March 4, 2004, 10:26:50
Job time : 83 secs

```

OM protein - protein search, using sw model

Run on: March 4, 2004, 09:18:40 ; Search time 48 Seconds  
(without alignments)  
291.810 Million cell updates/sec

Title: US-09-852-100B-2  
Perfect score: 1439  
Sequence: 1 MHILKGSPNVIPRAHGQKNT.....TRLTRLSITNETFRKTQLYP 269

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	167.5	11.6	573	1	YKK3_CAEEL	P34280 caenorhabdi
2	90	6.3	1324	1	VGL2_CVMA5	P11224 murine coro
3	89.5	6.2	515	1	EF1S_PORPU	P50257 porphyra pu
4	86.5	6.0	151	1	LCT2_MOUSE	O88803 mus musculu
5	85.5	5.9	348	1	ADH2_KLULA	P49383 kluyveromyc
6	85	5.9	338	1	LAMP_RAT	Q62813 rattus norv
7	83.5	5.8	487	1	Y346_MYCTU	O06297 mycobacteri
8	83	5.8	348	1	ADH1_KLUMA	Q07288 kluyveromyc
9	83	5.8	764	1	CFAB_HUMAN	P00751 homo sapien
10	82.5	5.7	338	1	LAMP_HUMAN	Q13449 homo sapien
11	81.5	5.7	493	1	GATA_RHIME	Q92qk7 rhizobium m
12	80.5	5.6	223	1	VG32_BPMD2	O64226 mycobacteri
13	80.5	5.6	455	1	ENT1_HUMAN	Q99808 homo sapien
14	80	5.6	328	1	IBP2_HUMAN	P18065 homo sapien
15	80	5.6	450	1	LIPP_PIG	P00591 sus scrofa
16	80	5.6	638	1	OAR1_LYMST	O77408 lymnaea sta
17	79.5	5.5	489	1	ANSP_MYCTU	O33261 mycobacteri
18	79.5	5.5	521	1	GATA_RHILO	Q98m95 rhizobium l

19	79.5	5.5	855	1	POLG_HRV3	Q82081	human rhino
20	79.5	5.5	910	1	AUXI_BOVIN	Q27974	bos taurus
21	79.5	5.5	961	1	GCSP_STRAW	Q827d7	streptomyce
22	79.5	5.5	3432	1	POLG_JAEV1	P27395	j genome po
23	79.5	5.5	3432	1	POLG_JAEV5	P19110	j genome po
24	79.5	5.5	3432	1	POLG_JAEVJ	P32886	j genome po
25	79.5	5.5	4367	1	DYHC_NEUCR	P45443	neurospora
26	79	5.5	468	1	LIP2_MOUSE	P17892	mus musculu
27	79	5.5	525	1	SUT_SPIOL	Q03411	spinacia ol
28	79	5.5	797	1	S6A5_HUMAN	Q9y345	homo sapien
29	79	5.5	1455	1	AIP1_HUMAN	Q86ul8	homo sapien
30	78.5	5.5	245	1	CRS3_HORSE	O19010	equus cabal
31	78.5	5.5	398	1	MUB1_XENLA	P38565	xenopus lae
32	78.5	5.5	5376	1	ZAN_MOUSE	O88799	mus musculu
33	78	5.4	557	1	YP85_MYCTU	Q50636	mycobacteri
34	78	5.4	557	1	YQ16_MYCBO	P59984	mycobacteri
35	77.5	5.4	172	1	K131_HUMAN	Q8iuc0	homo sapien
36	77.5	5.4	374	1	AROC_METTH	O26843	methanobact
37	77.5	5.4	2179	1	POLG_HRV14	P03303	h genome po
38	77	5.4	365	1	IDI2_METMA	Q8pw37	methanosarc
39	77	5.4	764	1	CFAB_PANTR	Q864w0	pan troglod
40	77	5.4	1302	1	NRG_DROME	P20241	drosophila
41	76.5	5.3	772	1	PMIP_YEAST	P35999	saccharomyc
42	76	5.3	379	1	METX_NEIMB	Q9jzq5	neisseria m
43	76	5.3	467	1	LIP1_HUMAN	P54315	homo sapien
44	76	5.3	484	1	GATA_PSEAE	Q9hvt8	pseudomonas
45	76	5.3	501	1	GATA_STRAW	Q82jl0	streptomyce

#### ALIGNMENTS

##### RESULT 1

##### YKK3\_CAEEL

ID YKK3\_CAEEL STANDARD; PRT; 573 AA.

AC P34280;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Hypothetical GTP-binding protein C02F5.3 in chromosome III.

GN C02F5.3.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RX MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,

RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,

RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,

RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,

RA Johnston L., Jones M., Kershaw J., Kirsten J., Laisstter N.,

RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,

RA Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R.,

RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
RA Wohldman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans."  
RL Nature 368:32-38(1994).  
CC -!- SIMILARITY: Belongs to the GTP1 / OBG family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

DR      EMBL; L14745; AAA27918.1; -.
DR      PIR; S44605; S44605.
DR      WormPep; C02F5.3; CE00039.
DR      InterPro; IPR006074; GTP1/OBG_dom.
DR      InterPro; IPR006073; GTP1_OBG.
DR      InterPro; IPR006169; GTP1_OBG_sub.
DR      InterPro; IPR005225; Small_GTP.
DR      InterPro; IPR004095; TGS_dom.
DR      InterPro; IPR007829; TM2.
DR      Pfam; PF01018; GTP1_OBG; 1.
DR      Pfam; PF02824; TGS; 1.
DR      Pfam; PF05154; TM2; 1.
DR      PRINTS; PR00326; GTP1OBG.
DR      TIGRFAMS; TIGR00231; small_GTP; 1.
DR      PROSITE; PS00905; GTP1_OBG; 1.
KW      Hypothetical protein; GTP-binding.
FT      NP_BIND          69          76          GTP (
FT      NP_BIND          115         119          GTP (
FT      NP_BIND          246         249          GTP (
SQ      SEQUENCE      573 AA;  64299 MW;  BA

```

Query Match 11.6%; Score 167.5; DB 1; Length 573;  
Best Local Similarity 27.9%; Pred. No. 1.7e-07;  
Matches 50; Conservative 23; Mismatches 57; Indels 49; Gaps 5;

Qy	90	VTTGPWGAVATSAGGEESLKCEDLKVGQYICKDP-----K	124
		:       :   :   :	
Db	415	VSTNPLGPV-----VECRFLENSFILCEDPVPLYGPGQTGQQPANESFRNEGKCLK	465
Qy	125	INDATQEPVNCNTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNVNGYSYK	184
		:                           : :	
Db	466	MGGYRAEDVEFTN---VKCRVLPCIEC---HGPR-----FTKSTPCIIYNHGYFL	510
Qy	185	VAVALSLFLGWLGLADRFYLGYPALGLLKFCVTGFCGIGSLIDFILISMQIVGPSDGSSY	243
		:   :     :         : :     :   :   :   :   :	
Db	511	TTLLYSIFLGVVAVDRFCLGYSAMAVGKMLTGGGFIWIVDIFLLVLGVLGPADDSSW	569

RESULT 2  
VGL2\_CVMA5  
ID\_VGL2 CVMA5 STANDARD; PRT; 1324 AA.

AC P11224;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE E2 glycoprotein precursor (Spike glycoprotein) (Peplomer protein)  
 DE [Contains: Spike protein S1 (90B); Spike protein S2 (90A)].  
 GN S.  
 OS Murine coronavirus (strain A59) (MHV-A59) (Murine hepatitis virus).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
 OC Coronaviridae; Coronavirus.  
 OX NCBI\_TaxID=111142;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88072088; PubMed=2825419;  
 RA Luytjes W., Sturman L.S., Bredenbeek P.J., Charite J.,  
 RA van der Zeijst B.A.M., Horzinek M.C., Spaan W.J.M.;  
 RT "Primary structure of the glycoprotein E2 of coronavirus MHV-A59 and  
 RT identification of the trypsin cleavage site.";  
 RL Virology 161:479-487(1987).  
 CC -!- FUNCTION: THE PEPLIMER PROTEIN MEDIATES THE BINDING OF VIRIONS  
 CC TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION  
 CC AND IN SYNCYTIUM FORMATION.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M18379; AAA46455.1; -.  
 DR InterPro; IPR002552; Corona\_S2.  
 DR Pfam; PF01601; Corona\_S2; 1.  
 KW Glycoprotein; Envelope protein; Transmembrane; Signal.  
 FT SIGNAL 1 16  
 FT CHAIN 17 1324 E2 GLYCOPROTEIN.  
 FT CHAIN 17 717 SPIKE PROTEIN S1.  
 FT CHAIN 718 1324 SPIKE PROTEIN S2.  
 FT DOMAIN 17 1265 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1266 1286 POTENTIAL.  
 FT DOMAIN 1287 1324 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 1287 1304 CYS-RICH.  
 FT CARBOHYD 31 31 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 60 60 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 357 357 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 435 435 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 530 530 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 625 625 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 657 657 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 665 665 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 688 688 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 737 737 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 754 754 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 893 893 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1180 1180 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1190 1190 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1209 1209 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1225 1225 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1246 1246 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1324 AA; 146019 MW; E198EF8F0BCDBF0E CRC64;

Query Match 6.3%; Score 90; DB 1; Length 1324;

Best Local Similarity 23.6%; Pred. No. 5;

Matches 59; Conservative 27; Mismatches 94; Indels 70; Gaps 13;

Qy 25 TGLYPMRG-PFKNLALLP-----FSLPLLGGGSGSGEKVSVSKMAAAWPSGPSA-- 73  
 || ||: | |:|||| | | | | : : ||| :|  
 Db 66 TGYYPVDGSKFRNLALRG TNSVSLSWFQPPYLNQFNDGIFAK--VQNLKTSTPSGATAYF 123  
 Qy 74 PEAVTARLVGVLFVSVTTGPWGA VATSAGGEESLKCEDLVKGQY-ICKDPKINDATQEP 132  
 | | | | : :| |: | :: | || ||: |  
 Db 124 PTIVIGSLFGYTSY-TVVIEPYNGVIMAS-----VCQYTICQLP----- 161  
 Qy 133 VNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNE-VGFF----KPISCRNVNGYSYKVAV 187  
 : | | | ||: :||: || | :: |  
 Db 162 -----YTDCKPNTN-----GNKLIGFWHTDVKKPICVLKRNF T LNVNA 199  
 Qy 188 ALSLFLGWL GADRFYLG Y---PALGLLK FCTVGF CGIGSLIDFILISMQIVGPSDGSSYI 244  
 | : || | |: | || :: : | |: ||::  
 Db 200 DAFYFHFYQHGGTFYAYYADKPSATTFLFSVY----IGDILTQYYVLPFICNPTAGSTFA 255  
 Qy 245 IDYYGTRLTR 254  
 |: | | :  
 Db 256 PRYWVTPLVK 265

### RESULT 3

EF1S\_PORPU

ID EF1S\_PORPU STANDARD; PRT; 515 AA.

AC P50257;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE Elongation factor 1-alpha S (EF-1-alpha S) (Sporophyte-specific EF-1-alpha).

GN TEF-S.

OS Porphyra purpurea.

OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.

OX NCBI\_TaxID=2787;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Avonport;

RX MEDLINE=96309386; PubMed=8704161;

RA Liu Q.Y., Baldauf S.L., Reith M.E.;

RT "Elongation factor 1 alpha genes of the red alga Porphyra purpurea

include a novel, developmentally specialized variant.";

RL Plant Mol. Biol. 31:77-85(1996).

CC -!- FUNCTION: This protein promotes the GTP-dependent binding of

aminoacyl-tRNA to the A-site of ribosomes during protein

biosynthesis.

CC -!- SUBCELLULAR LOCATION: Cytoplasmic.





DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Leukocyte cell-derived chemotaxin 2 precursor (Chondromodulin II)  
 DE (ChM-II).  
 GN LECT2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANT VAL-129.  
 RC STRAIN=BALB/c; TISSUE=Liver;  
 RX MEDLINE=98382586; PubMed=9714793;  
 RA Yamagoe S., Watanabe T., Mizuno S., Suzuki K.;  
 RT "The mouse Lect2 gene: cloning of cDNA and genomic DNA, structural  
 RT characterization and chromosomal localization.";  
 RL Gene 216:171-178(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC STRAIN=Swiss Webster / NIH; TISSUE=Embryo, and Liver;  
 RX MEDLINE=99160594; PubMed=10050029;  
 RA Shukunami C., Kondo J., Wakai H., Takahashi K., Inoue H., Kamizono A.,  
 RA Hiraki Y.;  
 RT "Molecular cloning of mouse and bovine chondromodulin-II cDNAs and the  
 RT growth-promoting actions of bovine recombinant protein.";  
 RL J. Biochem. 125:436-442(1999).  
 CC -!- FUNCTION: Has a neutrophil chemotactic activity. Also a positive  
 CC regulator of chondrocyte proliferation.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1; Synonyms=LECT2;  
 CC IsoId=O88803-1; Sequence=Displayed;  
 CC Name=2; Synonyms=LECT2Q;  
 CC IsoId=O88803-2; Sequence=VSP\_003051;  
 CC -!- TISSUE SPECIFICITY: Highly expressed in liver and weakly in  
 CC testis. Not expressed in heart, brain, spleen, lung, skeletal  
 CC muscle and kidney.  
 CC -!- SIMILARITY: Belongs to the LECT2 / MIM-1 family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB009687; BAA33383.1; -.  
 DR EMBL; AB009688; BAA33384.1; -.  
 DR EMBL; AB009689; BAA33385.1; -.  
 DR EMBL; AB009689; BAA33386.1; -.  
 DR EMBL; AF035161; AAF13302.1; -.  
 DR MGD; MGI:1278342; Lect2.  
 DR InterPro; IPR008663; LECT2.  
 DR Pfam; PF05429; LECT2; 1.  
 KW Chemotaxis; Signal; Alternative splicing.  
 FT SIGNAL 1 18 BY SIMILARITY.

FT CHAIN 19 151 LEUKOCYTE CELL-DERIVED CHEMOTAXIN 2.  
 FT VARSPLIC 98 151 FCVKIFYIKPIKYKGSIKKGEKLGTLPLQKIYPGIQSHVH  
 FT VENCDDSSDPTAYL -> QRLQAHTTTLNVFTCYWDKIQIPR  
 FT PTRFLCQNFLH (in isoform 2).  
 FT /FTid=VSP\_003051.  
 FT VARIANT 129 129 I -> V.  
 SQ SEQUENCE 151 AA; 16405 MW; 18AF444046B7AE8E CRC64;

Query Match 6.0%; Score 86.5; DB 1; Length 151;  
 Best Local Similarity 24.8%; Pred. No. 0.75;  
 Matches 29; Conservative 9; Mismatches 52; Indels 27; Gaps 4;

Qy 78 TARLVGVLFVSVTTGPGAVATSAGGEESLKCEDLVKGQYICKDPKINDATQEPVNCTN 137  
 | | : | | | : | | : | | : | |  
 Db 4 TTILISAALLSSALAGPWANICASKSSNEIRTCDSYGCGQYSAQ-----RTQR----- 51  
 Qy 138 YTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNV-----NGYSYKV 185  
 | | : | | | | | | | : | : | :  
 Db 52 ---H---HPGVDVLCSDGSVVYAPFTGKIVGQEKPYRNKNAINDGIRLSGRGFCVKI 102

# RESULT 5

## ADH2\_KLULA

ID ADH2\_KLULA STANDARD; PRT; 348 AA.  
 AC P49383;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Alcohol dehydrogenase II (EC 1.1.1.1).  
 GN ADH2.  
 OS Kluyveromyces lactis (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
 OX NCBI\_TaxID=28985;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CBS 2359 / IFO 1267 / NRRL Y-1140 / WM37;  
 RX MEDLINE=92269769; PubMed=1588917;  
 RA Shain D.H., Salvadore C., Denis C.L.;  
 RT "Evolution of the alcohol dehydrogenase (ADH) genes in yeast:  
 RT characterization of a fourth ADH in Kluyveromyces lactis."  
 RL Mol. Gen. Genet. 232:479-488(1992).  
 CC -!- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +  
 CC NADH.  
 CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).  
 CC -!- SUBUNIT: Homotetramer (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase  
 CC family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC -----
DR EMBL; X64397; CAA45739.1; -.
DR PIR; S20911; S20911.
DR InterPro; IPR002328; ADH_zinc.
DR InterPro; IPR002085; Adh_zn_family.
DR Pfam; PF00107; ADH_zinc_N; 1.
DR PROSITE; PS00059; ADH_ZINC; 1.
KW Oxidoreductase; Zinc; Metal-binding; NAD; Multigene family.
FT METAL      44      44      ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL      67      67      ZINC 1 (CATALYTIC) (BY SIMILARITY).
FT METAL      98      98      ZINC 2 (BY SIMILARITY).
FT METAL     101     101      ZINC 2 (BY SIMILARITY).
FT METAL     104     104      ZINC 2 (BY SIMILARITY).
FT METAL     112     112      ZINC 2 (BY SIMILARITY).
FT METAL     154     154      ZINC 1 (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE   348 AA;  37097 MW;  F3B64AE1F520689C CRC64;

```

```

Query Match          5.9%;  Score 85.5;  DB 1;  Length 348;
Best Local Similarity 19.9%;  Pred. No. 2.5;
Matches    65;  Conservative    35;  Mismatches 111;  Indels 115;  Gaps 13;

```

```

QY      19 NTRRDG---TGLYPMRGPFKNLALLPFSLPLLGSGSGEKVSVSKMAAAW----- 67
      | : |   | : : | :   || ||| : || | : | | : :   |
Db      37 NVKYSGVCHTDLHAWKGDWP----LPTKLPLV-GGHEGAGVVVAMGENVKGWIIGDFAGI 91

QY      68 -----PSGPSAPEAVT 78
      | |   |
Db      92 KWLNGSCMSCEYCELSNESNCPDADLSGYTHDGSFQQYATADAVQAARI PKGTDLAEVAP 151

QY      79 ARLVGV-----LWFVSVTTGPGAVATSAGGEESLKCEDLKV----- 115
      ||   |   : | | | : : || || : |
Db     152 ILCAGVTVYKALKSADLKAGDWVAISGACGGLGSLAIQYAKAMGYRVLGIDTGAEKAKLF 211

QY     116 ----GQYICKDPKINDATQEPVNCTNYTAH-----VSCFPAPNITCKDSSGNETHFTGN 165
      | : |   |   : | | | || ||   || |   : | |   |
Db     212 KELGGEYFVDYAVSKDLIKEIVDATNGGAHGVINVSVSEFAI-----EQSTNYVRSNGT 265

QY     166 EVGFFKPISCRNVNGYSYKVAVALSLFLGWLG--AD-RFYLGYPALGLLKFC TVGFCGIG 222
      |   |   : :   : | : : | : : | || | | : | || :   : | :
Db     266 VVLVGLPRDAKCKSDVFTQVVKSVSIVGSYVGNRADTREALDFFARGLV-HAPIKIVGLS 324

QY     223 SLIDFI--LISMQIVGPSDGSSYIID 246
      | |   : :   : |||   | : |
Db     325 ELADVYDKMVKGEIVG-----RYVVD 345

```

# RESULT 6

## LAMP\_RAT

```

ID LAMP_RAT          STANDARD;          PRT;    338 AA.
AC Q62813;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Limbic system-associated membrane protein precursor (LSAMP).
GN LSAMP OR LAMP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 29-49.  
RC TISSUE=Hippocampus;  
RX MEDLINE=95374785; PubMed=7646886;  
RA Pimenta A.F., Zhukareva V., Barbe M.F., Reinoso B.S., Grimley C.,  
RA Henzel W., Fischer I., Levitt P.;  
RT "The limbic system-associated membrane protein is an Ig superfamily  
RT member that mediates selective neuronal growth and axon targeting.";  
RL Neuron 15:287-297(1995).

CC -!- FUNCTION: MEDIATES SELECTIVE NEURONAL GROWTH AND AXON TARGETING.  
CC CONTRIBUTES TO THE GUIDANCE OF DEVELOPING AXONS AND REMODELING OF  
CC MATURE CIRCUITS IN THE LIMBIC SYSTEM. ESSENTIAL FOR NORMAL GROWTH  
CC OF THE HYPOCAMPAL MOSSY FIBER PROJECTION.  
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
CC -!- TISSUE SPECIFICITY: Expressed mostly by neurons comprising limbic-  
CC associated cortical and subcortical regions that function in  
CC cognition, emotion, memory, and learning.  
CC -!- DEVELOPMENTAL STAGE: FIRST DETECTED AT E15-16, AT STAGE E20 IT IS  
CC DETECTED IN PRESUMPTIVE CORTEX, MEDIAL LIMBIC AREAS OF THE  
CC THALAMUS AND HYPOTHALAMUS. IN THE ADULT, IT IS FOUND IN  
CC HYPOTHALAMUS, PERIRHINAL CORTEX, AMYGDALA AND MEDIAL THALAMIC  
CC REGION.  
CC -!- SIMILARITY: Belongs to the immunoglobulin superfamily. IgLON  
CC family.  
CC -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; U31554; AAA86120.1; -.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003598; Ig\_c2.  
DR Pfam; PF00047; ig; 3.  
DR SMART; SM00408; IGc2; 2.  
DR PROSITE; PS50835; IG\_LIKE; 3.  
KW Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;  
KW Repeat; Signal; Lipoprotein.  
FT SIGNAL 1 28  
FT CHAIN 29 315 LIMBIC SYSTEM-ASSOCIATED MEMBRANE  
FT PROTEIN.  
FT PROPEP 316 338 REMOVED IN MATURE FORM (POTENTIAL).  
FT DOMAIN 29 122 IG-LIKE C2-TYPE 1.  
FT DOMAIN 132 214 IG-LIKE C2-TYPE 2.  
FT DOMAIN 219 304 IG-LIKE C2-TYPE 3.  
FT DISULFID 53 111 POTENTIAL.  
FT DISULFID 153 197 POTENTIAL.  
FT DISULFID 239 290 POTENTIAL.  
FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 66 66 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 279 279 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 315 315 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT LIPID 315 315 GPI-anchor amidated asparagine  
 FT (Potential).  
 SQ SEQUENCE 338 AA; 37324 MW; 0B76AFDD68A39BB6 CRC64;

Query Match 5.9%; Score 85; DB 1; Length 338;  
 Best Local Similarity 27.7%; Pred. No. 2.7;  
 Matches 36; Conservative 15; Mismatches 47; Indels 32; Gaps 7;

QY 101 SAGGEESLKCEDLKVG---QYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSS 156  
 : | : |||| | : : | : | | : | | : | : :  
 Db 230 TTGRQASLKCEASAVPAPDFEWYRDDTRINSANGLEIKSTE---GQSSLTVTNVT-EEHY 285  
 QY 157 GNETHFTGNEVG-----FFKPISCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALG 209  
 || | | : : | | : | | | : | | | | | |  
 Db 286 GNYTCVAANKLGVTNASLVLFVRPGSVRGING-SISLAVPL-----WL-----LA 328  
 QY 210 LLKFCTVGFC 219  
 | | : |  
 Db 329 ASLFCLLSKC 338

#### RESULT 7

Y346\_MYCTU

ID Y346\_MYCTU STANDARD; PRT; 487 AA.

AC 006297;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Hypothetical transport protein Rv0346c/MT0361/Mb0354c.

GN RV0346C OR MT0361 OR MTCY13E10.06C OR MB0354C.

OS Mycobacterium tuberculosis, and

OS Mycobacterium bovis.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI\_TaxID=1773, 1765;

RN [1]

RP SEQUENCE FROM N.A.

RC SPECIES=M.tuberculosis; STRAIN=H37Rv;

RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,

RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,

RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,

RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,

RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,

RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;

RT "Deciphering the biology of Mycobacterium tuberculosis from the

RT complete genome sequence.";

RL Nature 393:537-544(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;  
 RX MEDLINE=22206494; PubMed=12218036;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;  
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains.";  
 RL J. Bacteriol. 184:5479-5490(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=M.bovis; STRAIN=AF2122/97;  
 RX MEDLINE=22709107; PubMed=12788972;  
 RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,  
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsempe C., Simon S.,  
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,  
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;  
 RT "The complete genome sequence of Mycobacterium bovis.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).  
 CC -!- FUNCTION: Probable amino-acid or metabolite transport protein.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -!- SIMILARITY: Belongs to the amino acid permease family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z95324; CAB08578.1; -.  
 DR EMBL; AE006942; AAK44583.1; -.  
 DR EMBL; BX248335; CAD93217.1; -.  
 DR PIR; C70574; C70574.  
 DR TIGR; MT0361; -.  
 DR TubercuList; Rv0346c; -.  
 DR InterPro; IPR002293; AA/rel\_permease1.  
 DR InterPro; IPR004840; AAc\_permease.  
 DR InterPro; IPR004841; Permease\_region.  
 DR Pfam; PF00324; aa\_permeases; 1.  
 DR PROSITE; PS00218; AMINO\_ACID\_PERMEASE\_1; 1.  
 KW Hypothetical protein; Transport; Amino-acid transport; Transmembrane;  
 KW Complete proteome.  
 FT TRANSMEM 26 46 POTENTIAL.  
 FT TRANSMEM 50 70 POTENTIAL.  
 FT TRANSMEM 98 118 POTENTIAL.  
 FT TRANSMEM 133 153 POTENTIAL.  
 FT TRANSMEM 163 183 POTENTIAL.  
 FT TRANSMEM 214 234 POTENTIAL.  
 FT TRANSMEM 256 276 POTENTIAL.  
 FT TRANSMEM 290 310 POTENTIAL.  
 FT TRANSMEM 341 361 POTENTIAL.  
 FT TRANSMEM 369 389 POTENTIAL.  
 FT TRANSMEM 414 434 POTENTIAL.  
 FT TRANSMEM 440 460 POTENTIAL.

SQ SEQUENCE 487 AA; 52194 MW; 3572502DB6ACD987 CRC64;

Query Match 5.8%; Score 83.5; DB 1; Length 487;  
Best Local Similarity 26.5%; Pred. No. 5.8;  
Matches 27; Conservative 16; Mismatches 40; Indels 19; Gaps 5;

```
Qy      159 ETHFTGNEVGFFKPISCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCTVGF 218
          : | : | : | : : : | : | | | | | | | | | | | | | | | | | |
Db      8 DERLTREDTGYHKGLHSRQLQMIALGGAIGTGLFLG--AGGRLASAGPGL----FLVYGI 61

Qy      219 CGIGSLIDFILISMQIVG-----PSDGS--SYIIDYYGTRL 252
          ||| | : : : : | | | | | | : : | : :
Db      62 CGI-----FVFLILRALGELVLHRPSSGSFVSFYAREFYGEKV 98
```

RESULT 8

ADH1\_KLUMA

ID ADH1\_KLUMA STANDARD; PRT; 348 AA.  
AC Q07288;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Alcohol dehydrogenase 1 (EC 1.1.1.1).  
GN ADH1.  
OS Kluyveromyces marxianus (Yeast) (Kluyveromyces fragilis).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
OX NCBI\_TaxID=4911;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 12424;  
RX MEDLINE=93250057; PubMed=8485163;  
RA Ladriere J.M., Delcour J., Vandenhaute J.;  
RT "Sequence of a gene coding for a cytoplasmic alcohol dehydrogenase  
RT from Kluyveromyces marxianus ATCC 12424.";  
RL Biochim. Biophys. Acta 1173:99-101(1993).  
CC -!- CATALYTIC ACTIVITY: An alcohol + NAD(+) = an aldehyde or ketone +  
CC NADH.  
CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).  
CC -!- SUBUNIT: Homotetramer.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: Belongs to the zinc-containing alcohol dehydrogenase  
CC family.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X60224; CAA42785.1; -.  
DR PIR; S32521; S32521.  
DR InterPro; IPR002328; ADH\_zinc.  
DR InterPro; IPR002085; Adh\_zn\_family.  
DR Pfam; PF00107; ADH\_zinc\_N; 1.



DR PROSITE; PS00059; ADH\_ZINC; 1.  
 KW Oxidoreductase; Zinc; Metal-binding; NAD; Multigene family.  
 FT METAL 44 44 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 FT METAL 67 67 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 FT METAL 98 98 ZINC 2 (BY SIMILARITY).  
 FT METAL 101 101 ZINC 2 (BY SIMILARITY).  
 FT METAL 104 104 ZINC 2 (BY SIMILARITY).  
 FT METAL 112 112 ZINC 2 (BY SIMILARITY).  
 FT METAL 154 154 ZINC 1 (CATALYTIC) (BY SIMILARITY).  
 SQ SEQUENCE 348 AA; 37158 MW; A75D2EBE82E355BD CRC64;

Query Match 5.8%; Score 83; DB 1; Length 348;  
 Best Local Similarity 20.8%; Pred. No. 4.3;  
 Matches 65; Conservative 38; Mismatches 107; Indels 102; Gaps 14;

QY 19 NTRRDG---TGLYPMRGPFPKNLALLPFSPLLLGGGGSGSGEKVSVSKMAAAWPSG----- 70  
 | : | | : : | | | : | | : : | |  
 Db 37 NVKYSGVCHTDLHAWQGDWP----LDTKLPLV-GGHEGAGIVVAMGENVTGWEIGDYAGI 91  
 QY 71 -----PSAPEA-----VTARLV----- 82  
 | : | | | |  
 Db 92 KWLNGSCMSCEECELSNEPNCPKADLSGYTHDGSFQQYATADAVQAARI PKNVDLAEVAP 151  
 QY 83 ----GV-----LWFVSVTTGPWGAVATSAGGEESLKCEDLKV----- 115  
 | | : | | : : | | : |  
 Db 152 ILCAGVTVYKALKSAHIKAGDWVAISGACGGLGSLAIQYAKAMGYRVLGIDAGDEKAKLF 211  
 QY 116 ----GQYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNE--VGF 169  
 | : | | : | | | : : : | | | |  
 Db 212 KELGGEYFIDFTKTKDMVAEVI EATNGVAHAVINVS VSEAAISTSVLYTRNGTVVLVGL 271  
 QY 170 FKPISCRNVNGYSYKVAVALSLFLGWLG--AD-RFYLGYPALGLLK--FCTVGFCGIGSL 224  
 : | : : : | : : : : | | | : : | : | : : | :  
 Db 272 PRDAQCK--SDVFNQVVK SISIVGSYVGNRADTREALDFFSRGLVKAPIKILGLSELASV 329  
 QY 225 IDFILISMQIVG 236  
 | : : | | |  
 Db 330 YD-KMVKGQIVG 340

# RESULT 9

## CFAB\_HUMAN

ID CFAB\_HUMAN STANDARD; PRT; 764 AA.  
 AC P00751; O15006; Q29944; Q96HX6; Q9BTF5; Q9BX92;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Complement factor B precursor (EC 3.4.21.47) (C3/C5 convertase)  
 DE (Properdin factor B) (Glycine-rich beta glycoprotein) (GBG) (PBF2).  
 GN BF.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANTS ARG-28; GLN-28; GLN-32  
 RP AND SER-736.

RX MEDLINE=91065702; PubMed=2249879;  
 RA Davrinche C., Abbal M., Clerc A.;  
 RT "Molecular characterization of human complement factor B subtypes.";   
 RL Immunogenetics 32:309-312(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANTS ARG-28 AND GLN-32.  
 RC TISSUE=Liver;  
 RX MEDLINE=94237735; PubMed=8181962;  
 RA Mejia J.E., Jahn I., de la Salle H., Hauptmann G.;  
 RT "Human factor B. Complete cDNA sequence of the BF\*S allele.";   
 RL Hum. Immunol. 39:49-53(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANTS ARG-28 AND GLN-32.  
 RC TISSUE=Liver;  
 RX MEDLINE=94041399; PubMed=8225386;  
 RA Schwaeble W., Luettig B., Sokolowski T., Estaller C., Weiss E.H.,  
 RA Meyer Zum Bueschenfelde K.-H., Whaley K., Dippold W.;  
 RT "Human complement factor B: functional properties of a recombinant  
 RT zymogen of the alternative activation pathway convertase.";   
 RL Immunobiology 188:221-232(1993).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANTS ARG-28 AND GLN-32.  
 RX MEDLINE=94067177; PubMed=8247029;  
 RA Horiuchi T., Kim S., Matsumoto M., Watanabe I., Fujita S.,  
 RA Volanakis J.E.;  
 RT "Human complement factor B: cDNA cloning, nucleotide sequencing,  
 RT phenotypic conversion by site-directed mutagenesis and expression.";   
 RL Mol. Immunol. 30:1587-1592(1993).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Rowen L., Dankers C., Baskin D., Faust J., Loretz C., Ahearn M.E.,  
 RA Banta A., Swartzell S., Smith T.M., Spies T., Hood L.;  
 RT "Sequence determination of 300 kilobases of the human class III MHC  
 RT locus.";   
 RL Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RA Jaatinen T., Kanerva J., Poutanen K.E., Saarinen-Pihkala U.,  
 RA Lokki M.-L.;  
 RT "Expression and alternative splicing of human factor B gene in  
 RT leukemic mononuclear cells.";   
 RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.  
 RN [7]  
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANTS HIS-9; GLN-32; TRP-32;  
 RP SER-252; GLU-565 AND GLU-651.  
 RA Rieder M.J., Carrington D.P., Hastings N.C., Ahearn M.O.,  
 RA Kuldanek S.A., Rajkumar N., Toth E.J., Yi Q., Nickerson D.A.;  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.  
 RN [8]  
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANT TRP-32.  
 RC TISSUE=Colon;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [9]  
 RP SEQUENCE OF 26-764, PARTIAL SEQUENCE FROM N.A., AND CARBOHYDRATES.  
 RX MEDLINE=84161997; PubMed=6546754;  
 RA Mole J.E., Anderson J.K., Davison E.A., Woods D.E.;  
 RT "Complete primary structure for the zymogen of human complement  
 RT factor B.";  
 RL J. Biol. Chem. 259:3407-3412(1984).  
 RN [10]  
 RP SEQUENCE OF 260-764.  
 RX MEDLINE=83204002; PubMed=6342610;  
 RA Christie D.L., Gagnon J.;  
 RT "Amino acid sequence of the Bb fragment from complement Factor B.  
 RT Sequence of the major cyanogen bromide-cleavage peptide (CB-II) and  
 RT completion of the sequence of the Bb fragment.";  
 RL Biochem. J. 209:61-70(1983).  
 RN [11]  
 RP SEQUENCE OF 339-764 FROM N.A.  
 RX MEDLINE=83273641; PubMed=6308626;  
 RA Campbell R.D., Porter R.R.;  
 RT "Molecular cloning and characterization of the gene coding for human  
 RT complement protein factor B.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:4464-4468(1983).  
 RN [12]  
 RP SEQUENCE OF 467-595 AND 752-764 FROM N.A.  
 RX MEDLINE=83039428; PubMed=6957884;  
 RA Woods D.E., Markham A.F., Ricker A.T., Goldberger G., Colten H.R.;  
 RT "Isolation of cDNA clones for the human complement protein factor B,  
 RT a class III major histocompatibility complex gene product.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 79:5661-5665(1982).  
 RN [13]  
 RP SEQUENCE OF 16-259 FROM N.A.  
 RX MEDLINE=84158524; PubMed=6323161;  
 RA Morley B.J., Campbell R.D.;  
 RT "Internal homologies of the Ba fragment from human complement  
 RT component Factor B, a class III MHC antigen.";  
 RL EMBO J. 3:153-157(1984).  
 RN [14]  
 RP SEQUENCE OF 1-99 FROM N.A.  
 RC TISSUE=Blood;  
 RX MEDLINE=87102880; PubMed=3643061;  
 RA Wu L.C., Morley B.J., Campbell R.D.;  
 RT "Cell-specific expression of the human complement protein factor B

RT gene: evidence for the role of two distinct 5'-flanking elements.";  
 RL Cell 48:331-342(1987).  
 RN [15]  
 RP GLYCATION OF LYS-291.  
 RX MEDLINE=91174758; PubMed=2006911;  
 RA Niemann M.A., Bhowan A.S., Miller E.J.;  
 RT "The principal site of glycation of human complement factor B.";  
 RL Biochem. J. 274:473-480(1991).  
 CC -!- FUNCTION: Factor B which is part of the alternate pathway of the  
 CC complement system is cleaved by factor D into 2 fragments: Ba and  
 CC Bb. Bb, a serine protease, then combines with complement factor 3b  
 CC to generate the C3 or C5 convertase. It has also been implicated  
 CC in proliferation and differentiation of preactivated B  
 CC lymphocytes, rapid spreading of peripheral blood monocytes,  
 CC stimulation of lymphocyte blastogenesis and lysis of erythrocytes.  
 CC Ba inhibits the proliferation of preactivated B lymphocytes.  
 CC -!- CATALYTIC ACTIVITY: Cleaves C3 in the alpha-chain to yield C3a and  
 CC C3b. Cleaves C5 in the alpha-chain to yield C5a and C5b. Both  
 CC cleavages take place at the C-terminal of an arginine residue.  
 CC -!- SUBUNIT: Monomer.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=P00751-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=P00751-2; Sequence=VSP\_005380, VSP\_005381;  
 CC -!- POLYMORPHISM: Two major variants, F and S, and 2 minor variants,  
 CC as well as at least 14 very rare variants, have been identified.  
 CC -!- SIMILARITY: Belongs to peptidase family S1.  
 CC -!- SIMILARITY: Contains 3 Sushi (SCR) domains.  
 CC -!- SIMILARITY: Contains 1 VWFA domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X72875; CAA51389.1; -.  
 DR EMBL; S67310; AAD13989.1; -.  
 DR EMBL; L15702; AAA16820.1; -.  
 DR EMBL; X00284; CAA25077.1; -.  
 DR EMBL; AF019413; AAB67977.1; -.  
 DR EMBL; AF349679; AAK30167.1; -.  
 DR EMBL; AF551848; AAN71991.1; -.  
 DR EMBL; BC004143; AAH04143.1; -.  
 DR EMBL; BC007990; AAH07990.1; -.  
 DR EMBL; K01566; AAA36225.2; -.  
 DR EMBL; J00125; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; J00126; AAA36226.1; -.  
 DR EMBL; J00185; AAA36219.1; ALT\_SEQ.  
 DR EMBL; J00186; AAA36220.1; -.  
 DR EMBL; M15082; AAA59625.1; -.  
 DR PIR; S34075; BBHU.

```
DR HSSP; P20231; 1AAO.
DR MEROPS; S01.196; -.
DR SWISS-2DPAGE; P00751; HUMAN.
DR Siena-2DPAGE; P00751; -.
DR Genew; HGNC:1037; BF.
DR MIM; 138470; -.
DR GO; GO:0003811; F:complement activity; TAS.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001314; Peptidase_S1A.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00084; sushi; 3.
DR Pfam; PF00089; trypsin; 1.
```

Query Match 5.8%; Score 83; DB 1; Length 764;  
Best Local Similarity 24.1%; Pred. No. 11;  
Matches 49; Conservative 21; Mismatches 71; Indels 62; Gaps 12;

Qy	24	GTGLYPMRGPFKNLALLPFSPLPLGGG-----GSGSGEKVSV-----	60
		:       :                   :	
Db	2	GSNLSP-----QLCLMPFILGLLSGGVTTTPWSLAR PQGSCSLEGVEIKGGSFRLLQEG	55
Qy	61	SKMAAAWPSG--PSAPEAVTARLVGVLWFVSVTTGPGAVATS-----AGGEESLKC--	110
		:         :     :     :   :	
Db	56	QALEYVCPSGFYYPVQTRTCR-----STGSWSTLKTQDQKTVRKAECRAIH CPR	105
Qy	111	-EDLKV GQYICKDPKINDATQEPVNC-TNYTAHVSCFPAPNITCKDSS--GNETHFTGNE	166
		:   :   :   : : :           : : :	
Db	106	PHDFENG EYWPRSPYYNVSDEISFHCYDGYTLR GSA---NRTCQVNGRWSGQT AICDNG	161
Qy	167	VGFFK---PISCRNVNGYSYKV	185
		:           : :	
Db	162	AGYCSNPGIPIGTRKV-GSQYRL	183

RESULT 10

LAMP HUMAN

```

ID      LAMP_HUMAN          STANDARD;          PRT;      338 AA.
AC      Q13449;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Limbic system-associated membrane protein precursor (LSAMP).
GN      LSAMP OR LAMP.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=96235133; PubMed=8666243;
RA      Pimenta A.F., Fischer I., Levitt P.;
RT      "cDNA cloning and structural analysis of the human limbic-system-
RT      associated membrane protein (LAMP).";
RL      Gene 170:189-195(1996).
CC      -!- FUNCTION: MEDIATES SELECTIVE NEURONAL GROWTH AND AXON TARGETING.

```

```

CC      CONTRIBUTES TO THE GUIDANCE OF DEVELOPING AXONS AND REMODELING OF
CC      MATURE CIRCUITS IN THE LIMBIC SYSTEM. ESSENTIAL FOR NORMAL GROWTH
CC      OF THE HYPOCAMPAL MOSSY FIBER PROJECTION (BY SIMILARITY).
CC      -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC      -!- TISSUE SPECIFICITY: Expressed on limbic neurons and fiber tracts
CC      as well as in single layers of the superior colliculus, spinal
CC      chord and cerebellum.
CC      -!- SIMILARITY: Belongs to the immunoglobulin superfamily. IgLON
CC      family.
CC      -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U41901; AAC50569.1; -.
DR      PIR; JC4776; JC4776.
DR      Genew; HGNC:6705; LSAMP.
DR      MIM; 603241; -.
DR      GO; GO:0007399; P:neurogenesis; TAS.
DR      InterPro; IPR007110; Ig-like.
DR      InterPro; IPR003598; Ig_c2.
DR      Pfam; PF00047; ig; 3.
DR      SMART; SM00408; IGc2; 2.
DR      PROSITE; PS50835; IG_LIKE; 3.
KW      Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW      Repeat; Signal; Lipoprotein.
FT      SIGNAL          1      28      POTENTIAL.
FT      CHAIN           29     315     LIMBIC SYSTEM-ASSOCIATED MEMBRANE
FT                                     PROTEIN.
FT      PROPEP          316     338     REMOVED IN MATURE FORM (POTENTIAL).
FT      DOMAIN           29     122     IG-LIKE C2-TYPE 1.
FT      DOMAIN          132     214     IG-LIKE C2-TYPE 2.
FT      DOMAIN          219     304     IG-LIKE C2-TYPE 3.
FT      DISULFID         53     111     POTENTIAL.
FT      DISULFID        153     197     POTENTIAL.
FT      DISULFID        239     290     POTENTIAL.
FT      CARBOHYD         40      40     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD         66      66     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD        136     136     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD        148     148     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD        279     279     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD        287     287     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD        300     300     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD        315     315     N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      LIPID           315     315     GPI-anchor amidated asparagine
FT                                     (Potential).
SQ      SEQUENCE      338 AA;  37308 MW;  03455F286DF5D92F CRC64;

```

Query Match 5.7%; Score 82.5; DB 1; Length 338;

Best Local Similarity 29.6%; Pred. No. 4.6;

Matches 37; Conservative 14; Mismatches 47; Indels 27; Gaps 7;

Qy 101 SAGGEESLKCEDLKVG----QYICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSS 156  
 : | : |||| | : : | :|| | : | | :| : :  
 Db 230 TTGRQASLKCEASAVPAPDFEWYRDDTRINSANGLEIKSTE---GQSSLTVTNVT-EEHY 285

Qy 157 GNETHFTGNEVG-----FFKPISCRNVNGYSYKVAVALSLFLGWLGADRFYLGYPALG 209  
 || | | : : | | : | | | : || | : || | | |  
 Db 286 GNYTCVAANKLGVTNASLVLFRPGSVRGING-SISLAVPL-----WL-----LAASLLC 333

Qy 210 LLKFC 214  
 || |  
 Db 334 LLSKC 338

RESULT 11

GATA\_RHIME

ID GATA\_RHIME STANDARD; PRT; 493 AA.  
 AC Q92QK7;  
 DT 15-MAR-2004 (Rel. 43, Created)  
 DT 15-MAR-2004 (Rel. 43, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Glutamyl-tRNA(Gln) amidotransferase subunit A (EC 6.3.5.-) (Glu-ADT  
 DE subunit A).  
 GN GATA OR R01312 OR SMC01352.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.  
 OX NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RX MEDLINE=21396507; PubMed=11481430;  
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,  
 RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,  
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,  
 RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,  
 RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;  
 RT "Analysis of the chromosome sequence of the legume symbiont  
 RT Sinorhizobium meliloti strain 1021.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
 CC -!- FUNCTION: Furnishes a means for formation of correctly charged  
 CC Gln-tRNA(Gln) through the transamidation of misacylated Glu-  
 CC tRNA(Gln) in organisms which lack glutaminyl-tRNA synthetase. The  
 CC reaction takes place in the presence of glutamine and ATP through  
 CC an activated gamma-phospho-Glu-tRNA(Gln) (By similarity).  
 CC -!- CATALYTIC ACTIVITY: ATP + L-glutamyl-tRNA(Gln) + L-glutamine = ADP  
 CC + phosphate + L-glutaminyl-tRNA(Gln) + L-glutamate.  
 CC -!- SUBUNIT: Heterotrimer of A, B and C subunits (By similarity).  
 CC -!- SIMILARITY: Belongs to the amidase family.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AL591786; CAC45891.1; -.  
 DR HAMAP; MF\_00120; -; 1.  
 DR InterPro; IPR000120; Amidase.  
 DR InterPro; IPR004412; GatA.  
 DR Pfam; PF01425; Amidase; 1.  
 DR TIGRFAMs; TIGR00132; gatA; 1.  
 DR PROSITE; PS00571; AMIDASES; 1.  
 KW Protein biosynthesis; Ligase; Complete proteome.  
 SQ SEQUENCE 493 AA; 52654 MW; 1B7D595A197EF425 CRC64;

Query Match 5.7%; Score 81.5; DB 1; Length 493;  
 Best Local Similarity 21.1%; Pred. No. 8.9;  
 Matches 60; Conservative 38; Mismatches 116; Indels 71; Gaps 15;

QY 3 ILKGSPNVIPRAHGQKNTRRDGTGLY-PMRGPFK----NLALLPFSLPLLGGGGSGSGEK 57  
 :: | | : | | | | | | | : : : | | : | | |  
 Db 117 VMLGKLNLMDEFAMGSSNE---TSYYGPVKNPWRAGSNLDLVP-----GGSSGGSAAA 166  
 QY 58 VSVSKMAAAWPSGPSAPEAVTARLVGVLFVSVTTG---PWGAVATSAGGEESLK-CEDL 113  
 | : | | : | | : : | | | | : : : | :  
 Db 167 VAARLCAGATATDTGGSIRQPAFTGTVG-IKPTYGRCSRWGVVAFASSLDQAGPIARDV 225  
 QY 114 KVGQYICK-----DPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVG 168  
 : : | | | | : : | : | : | : |  
 Db 226 RDAAILLKSMASIDPK--DTTSVDLPVPDYEAAG-----QSIKGMRI 267  
 QY 169 FFKPISCRNVNGYSYKVAVALSLFLGLWL---GADRFYLGYPALGLLKFCCTVGF CGIGSLI 225  
 | : | : : : | | | : |  
 Db 268 IPKEY---RVDGMPEDIEALWQQGIAWLRDAGAEIVDISLPH-----T 307  
 QY 226 DFILISMQIVGPSDGSSYIIDY---YGTRLTRLSTNETFRKTQ 266  
 : | : | | : : | | : | | : | : : | :  
 Db 308 KYALPAYYIVAPAEASSNLARYDGVRYGLRVDGKDII-DMYEKTR 351

# RESULT 12

VG32\_BPM2

ID VG32\_BPM2 STANDARD; PRT; 223 AA.

AC 064226;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE Gene 32 protein (GP32).

GN 32.

OS Mycobacteriophage D29.

OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.

OX NCBI\_TaxID=28369;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98300335; PubMed=9636706;

RA Ford M.E., Sarkis G.J., Belanger A.E., Hendrix R.W., Hatfull G.F.;

RT "Genome structure of mycobacteriophage D29: implications for phage evolution.";

RL J. Mol. Biol. 279:143-164(1998).

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -



CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AF022214; AAC18473.1; -.

DR PIR; F72803; F72803.

SQ SEQUENCE 223 AA; 21822 MW; 33CD0DC310038AD4 CRC64;

Query Match 5.6%; Score 80.5; DB 1; Length 223;

Best Local Similarity 30.7%; Pred. No. 4.2;

Matches 27; Conservative 8; Mismatches 36; Indels 17; Gaps 3;

```
Qy      29 PMRGPFKNLALLPFSLP-----LLGGGGSGSGEKVSVSKMAAAWPSGPSAPEAVTA 79
      | : | : : : | | | | | | | | | | | | | | | | |
Db      37 PVLTPVTAVGAYTYNIPQAQAEFIDVILLGAGGGGQG-----MGSATAWGQGGFGGSWVTA 91

Qy      80 RL---VGVLWFFVSVTTGPGWGA VATSAGG 104
      | | : | | : | | | | : |
Db      92 TLRRGVDIPWAVTQITGVIGAGGTAGPG 119
```

#### RESULT 13

##### ENT1\_HUMAN

ID ENT1\_HUMAN STANDARD; PRT; 455 AA.

AC Q99808; Q9UJY2;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Equilibrative nucleoside transporter 1 (Equilibrative

DE nitrobenzylmercaptapurine riboside-sensitive nucleoside transporter)

DE (Equilibrative NBMPR-sensitive nucleoside transporter) (Nucleoside

DE transporter, es-type).

GN SLC29A1 OR ENT1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-21.

RC TISSUE=Placenta;

RX MEDLINE=97140266; PubMed=8986748;

RA Griffiths M., Beaumont N., Yao S.Y.M., Sundaram M., Boumah C.E.,

RA Davies A., Kwong F.Y.P., Coe I., Cass C.E., Young J.D., Baldwin S.A.;

RT "Cloning of a human nucleoside transporter implicated in the cellular

RT uptake of adenosine and chemotherapeutic drugs.";

RL Nat. Med. 3:89-93(1997).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Jejunum, and Small intestine;

RX MEDLINE=20216090; PubMed=10755314;

RA Lum P.Y., Ngo L.Y., Bakken A.H., Unadkat J.D.;

RT "Human intestinal es nucleoside transporter: molecular

RT characterization and nucleoside inhibitory profiles.";

RL Cancer Chemother. Pharmacol. 45:273-278(2000).

RN [3]

RP SEQUENCE FROM N.A.  
 RA Graham K.A., Coe I.R., Carpenter P., Baldwin S.A., Young J.D.,  
 RA Cass C.E.;  
 RT "Genomic sequence of the human equilibrative nucleoside transporter 1  
 RT (hENT1).";  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22272396; PubMed=12384580;  
 RA Sankar N., Machado J., Abdulla P., Hilliker A.J., Coe I.R.;  
 RT "Comparative genomic analysis of equilibrative nucleoside transporters  
 RT suggests conserved protein structure despite limited sequence  
 RT identity.";  
 RL Nucleic Acids Res. 30:4339-4350(2002).  
 RN [5]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Colon, and Muscle;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnierch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -!- FUNCTION: MEDIATES BOTH INFLUX AND EFFLUX OF NUCLEOSIDES ACROSS  
 CC THE MEMBRANE (EQUILIBRATIVE TRANSPORTER). IT IS SENSITIVE (ES) TO  
 CC LOW CONCENTRATIONS OF THE INHIBITOR NITROBENZYL MERCAPTOPYRINE  
 CC RIBOSIDE (NBMPR) AND IS SODIUM-INDEPENDENT. IT HAS A HIGHER  
 CC AFFINITY FOR ADENOSINE. INHIBITED BY DIPYRIDAMOLE AND DILAZEP  
 CC (ANTICANCER CHEMOTHERAPEUTICS DRUGS).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, MAMMARY GLAND,  
 CC ERYTHROCYTES AND PLACENTA, AND ALSO IN FETAL LIVER AND SPLEEN.  
 CC -!- PTM: Glycosylated.  
 CC -!- SIMILARITY: BELONGS TO THE SLC29A FAMILY OF TRANSPORTERS.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC -----
DR EMBL; U81375; AAC51103.1; -.
DR EMBL; AF079117; AAC62495.1; -.
DR EMBL; AF190884; AAF02777.1; -.
DR EMBL; AF495730; AAM11785.1; -.
DR EMBL; BC001382; AAH01382.1; -.
DR EMBL; BC008954; AAH08954.1; -.
DR Genew; HGNC:11003; SLC29A1.
DR MIM; 602193; -.
DR GO; GO:0005887; C:integral to plasma membrane; TAS.
DR GO; GO:0005624; C:membrane fraction; TAS.
DR GO; GO:0005337; F:nucleoside transporter activity; TAS.
DR GO; GO:0006139; P:nucleobase, nucleoside, nucleotide and nucl. . .; TAS.
DR GO; GO:0015858; P:nucleoside transport; TAS.
DR InterPro; IPR002259; DER/eqnu_transpt.
DR Pfam; PF01733; Nucleoside_tran; 1.
DR PRINTS; PR01130; DERENTRNSPRT.
DR ProDom; PD005103; DER/eqnu_transpt; 1.
DR TIGRFAMs; TIGR00939; 2a57; 1.
KW Transmembrane; Transport; Glycoprotein.
FT INIT_MET      0      0
FT DOMAIN        1     11      CYTOPLASMIC (POTENTIAL).
FT TRANSMEM      12     28      POTENTIAL.
FT DOMAIN        29     81      EXTRACELLULAR (POTENTIAL).
FT TRANSMEM      82    106      POTENTIAL.
FT DOMAIN       107    110      CYTOPLASMIC (POTENTIAL).
FT TRANSMEM     111    129      POTENTIAL.
FT DOMAIN       130    137      EXTRACELLULAR (POTENTIAL).
FT TRANSMEM     138    156      POTENTIAL.
FT DOMAIN       157    173      CYTOPLASMIC (POTENTIAL).
FT TRANSMEM     174    198      POTENTIAL.
FT DOMAIN       199    205      EXTRACELLULAR (POTENTIAL).
FT TRANSMEM     206    226      POTENTIAL.
FT DOMAIN       227    290      CYTOPLASMIC (POTENTIAL).
FT TRANSMEM     291    310      POTENTIAL.
FT DOMAIN       311    322      EXTRACELLULAR (POTENTIAL).
FT TRANSMEM     323    341      POTENTIAL.
FT DOMAIN       342    358      CYTOPLASMIC (POTENTIAL).
FT TRANSMEM     359    377      POTENTIAL.
FT DOMAIN       378    392      EXTRACELLULAR (POTENTIAL).
FT TRANSMEM     393    412      POTENTIAL.
FT DOMAIN       413    430      CYTOPLASMIC (POTENTIAL).
FT TRANSMEM     431    451      POTENTIAL.
FT DOMAIN       452    455      EXTRACELLULAR (POTENTIAL).
FT CARBOHYD      47     47      N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE      455 AA; 50088 MW; 9098E95E26515850 CRC64;

```

```

Query Match          5.6%; Score 80.5; DB 1; Length 455;
Best Local Similarity 22.6%; Pred. No. 9.9;
Matches 52; Conservative 39; Mismatches 82; Indels 57; Gaps 14;

```

```

Qy      30 MRGPFFKNLA-LLP--FSLPLLGSGSGSGEKVSVSKMAAAWPSGPSAPEA-----VTARL 81
      ::|  || |||  ::|::|  :|  ||:|  |  ||  |:  :||
Db      156 LQGSFLGLAGLLPASYTAPIMSGQGL-AGFFASVA-MICAIASGSELSESAFGYFITACA 213

Qy      82 VGVLWFVSVTTGP-----WGAVATSAGGEESLKCEDLKVGYICKDPK-----IND 127
      |:|  :  |  :  :  ||:|  :  :  |:  :|:  ::

```

```

Db      214 VIILTIICYLGLPRLEFYRYYYQQLKLEGPGEQETKLDLISKGE----EPRAGKEESGVSV 269
Qy      128 ATQEPVN-----CTNYTAHVSCFPAPNITCKDS-SGNETHFTGNEV 167
      :  :| |           |  :| : ||| : | | :| : |
Db      270 SNSQPTNESHSIKAILKNISVLAFSVCFIFTITIGMFPAVTVEVKSSIAGSSTW-----E 324
Qy      168 GFFKPISC-RNVNGYSYKVAVALSLFLGWLGADRFYLGYPALGLLKFCTV 216
      :| | :| |   | : :   : :| : | | :| | :| | : |
Db      325 RYFIPVSCFLT FNIFDWLGRSLTAVFM-WPGKDSRWL--PSLVLARLVFV 371

```

# RESULT 14

## IBP2\_HUMAN

```

ID      IBP2_HUMAN      STANDARD;      PRT;      328 AA.
AC      P18065; Q14619;
DT      01-NOV-1990 (Rel. 16, Created)
DT      01-NOV-1990 (Rel. 16, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Insulin-like growth factor binding protein 2 precursor (IGFBP-2)
DE      (IBP-2) (IGF-binding protein 2).
GN      IGFBP2 OR BP2.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Retina;
RX      MEDLINE=91293227; PubMed=1712312;
RA      Agarwal N., Hsieh C.L., Sills D., Swaroop M., Desai B., Francke U.,
RA      Swaroop A.;
RT      "Sequence analysis, expression and chromosomal localization of a
RT      gene, isolated from a subtracted human retina cDNA library, that
RT      encodes an insulin-like growth factor binding protein (IGFBP2).";
RL      Exp. Eye Res. 52:549-561(1991).
RN      [2]
RP      SEQUENCE FROM N.A., AND SEQUENCE OF 40-77.
RX      MEDLINE=90368661; PubMed=1697583;
RA      Zapf J., Kiefer M., Merryweather J., Masiarz F., Bauer D., Born W.,
RA      Fischer J.A., Foresch E.R.;
RT      "Isolation from adult human serum of four insulin-like growth factor
RT      (IGF) binding proteins and molecular cloning of one of them that is
RT      increased by IGF I administration and in extrapancreatic tumor
RT      hypoglycemia.";
RL      J. Biol. Chem. 265:14892-14898(1990).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Fetal liver;
RX      MEDLINE=90060007; PubMed=2479552;
RA      Binkert C., Landwehr J., Mary J.L., Schwander J., Heinrich G.;
RT      "Cloning, sequence analysis and expression of a cDNA encoding a novel
RT      insulin-like growth factor binding protein (IGFBP-2).";
RL      EMBO J. 8:2497-2502(1989).
RN      [4]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Placenta;
RX      MEDLINE=91248211; PubMed=1710112;

```

RA Ehrenborg E., Vilhelmsdotter S., Bajalica S., Larsson C., Sterm I.,  
RA Koch J., Brondum-Nielsen K., Luthman H.;  
RT "Structure and localization of the human insulin-like growth factor-  
RT binding protein 2 gene.";  
RL Biochem. Biophys. Res. Commun. 176:1250-1255(1991).  
RN [5]

RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=92293159; PubMed=1376411;  
RA Binkert C., Margot J.B., Landwehr J., Heinrich G., Schwander J.;  
RT "Structure of the human insulin-like growth factor binding protein-2  
RT gene.";  
RL Mol. Endocrinol. 6:826-836(1992).  
RN [6]

RP SEQUENCE FROM N.A.  
RC TISSUE=Brain, and Uterus;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnierch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -!- FUNCTION: IGF-binding proteins prolong the half-life of the IGFs  
CC and have been shown to either inhibit or stimulate the growth  
CC promoting effects of the IGFs on cell culture. They alter the  
CC interaction of IGFs with their cell surface receptors.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- MISCELLANEOUS: Binds IGF-II more than IGF-I.  
CC -!- SIMILARITY: Contains 1 IGFBP domain.  
CC -!- SIMILARITY: Contains 1 thyroglobulin type-I domain.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; S37730; AAB22308.1; -.  
DR EMBL; S37712; AAB22308.1; JOINED.  
DR EMBL; S37722; AAB22308.1; JOINED.

DR EMBL; S37726; AAB22308.1; JOINED.  
 DR EMBL; M35410; AAA03246.1; -.  
 DR EMBL; X16302; CAA34373.1; -.  
 DR EMBL; M69241; AAA36048.1; -.  
 DR EMBL; M69237; AAA36048.1; JOINED.  
 DR EMBL; M69239; AAA36048.1; JOINED.  
 DR EMBL; M69240; AAA36048.1; JOINED.  
 DR EMBL; A09809; CAA00862.1; -.  
 DR EMBL; BC004312; AAH04312.1; -.  
 DR EMBL; BC009902; AAH09902.1; -.  
 DR EMBL; BC012769; AAH12769.1; -.  
 DR PIR; A41927; A41927.  
 DR HSSP; P24593; 1BOE.  
 DR Genew; HGNC:5471; IGFBP2.  
 DR MIM; 146731; -.  
 DR InterPro; IPR009030; Grow\_fac\_recep.  
 DR InterPro; IPR000867; Insl\_gro\_fac\_pr.  
 DR InterPro; IPR000716; Thyroglobulin\_1.  
 DR Pfam; PF00219; IGFBP; 1.  
 DR Pfam; PF00086; thyroglobulin\_1; 1.  
 DR PIRSF; PIRSF001969; IGFBP1-6; 1.  
 DR SMART; SM00121; IB; 1.  
 DR SMART; SM00211; TY; 1.  
 DR PROSITE; PS00222; IGF\_BINDING; 1.  
 DR PROSITE; PS00484; THYROGLOBULIN\_1; 1.  
 KW Growth factor binding; Signal.  
 FT SIGNAL 1 39  
 FT CHAIN 40 328 INSULIN-LIKE GROWTH FACTOR BINDING  
 FT PROTEIN 2.  
 FT DOMAIN 260 309 THYROGLOBULIN TYPE-I.  
 FT SITE 304 306 CELL ATTACHMENT SITE.  
 FT CONFLICT 60 60 P -> R (IN REF. 4).  
 FT CONFLICT 320 320 R -> C (IN REF. 3).  
 FT CONFLICT 323 323 H -> D (IN REF. 4).  
 SQ SEQUENCE 328 AA; 35137 MW; 4E6BDF6D805C8853 CRC64;

Query Match 5.6%; Score 80; DB 1; Length 328;  
 Best Local Similarity 27.8%; Pred. No. 7.4;  
 Matches 40; Conservative 10; Mismatches 48; Indels 46; Gaps 8;

QY 22 RDGTGLYPMRGPFKNLALLPFSLLPGGGSGSGEKVSV-----SKMAAAWPSGPS 72  
 | | | : | | | | | | | : | : | | |  
 Db 4 RVGCPALPLPPP-PLLPLPLLLLLLLGASGGGGGARAELVFRCPCTPERLAACGPP-PV 61  
 QY 73 APEAVTARLVGVLFVSVTTGPWGAVATSAGGEESLKCEDLVKGQYICKDPKIN----DA 128  
 || | || | | : | : | : | |  
 Db 62 APPA-----AVAAVAGGAR-MPCAEL-----VREPGCGCCSVCA 94  
 QY 129 TQEPVNCTNYTAH----VSCFPAP 148  
 | | | : | : |  
 Db 95 RLEGEACGVYTPRCGQGLRCYPHP 118

# RESULT 15

LIPP\_PIG

ID LIPP\_PIG STANDARD; PRT; 450 AA.  
 AC P00591;

DT 21-JUL-1986 (Rel. 01, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Triacylglycerol lipase, pancreatic (EC 3.1.1.3) (Pancreatic lipase)  
 DE (PL).  
 GN PNLIP.  
 OS *Sus scrofa* (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE OF 308-449.  
 RX MEDLINE=82113655; PubMed=7326260;  
 RA de Caro J.D., Boudouard M., Bonicel J.J., Guidoni A.A., Desnuelle P.,  
 RA Roverly M.;  
 RT "Porcine pancreatic lipase. Completion of the primary structure.";  
 RL Biochim. Biophys. Acta 671:129-138(1981).  
 RN [2]  
 RP SEQUENCE OF 1-234, AND CARBOHYDRATE-LINKAGE SITE.  
 RX MEDLINE=79236335; PubMed=380992;  
 RA Bianchetta J.D., Bidaud J., Guidoni A.A., Bonicel J.J., Roverly M.;  
 RT "Porcine pancreatic lipase. Sequence of the first 234 amino acids of  
 RT the peptide chain.";  
 RL Eur. J. Biochem. 97:395-405(1979).  
 RN [3]  
 RP SEQUENCE OF 235-307.  
 RX MEDLINE=80088446; PubMed=518929;  
 RA Guidoni A.A., Bonicel J.J., Bianchetta J.D., Roverly M.;  
 RT "Porcine pancreatic lipase. Sequence between the 235th and 307th  
 RT amino acids.";  
 RL Biochimie 61:841-845(1979).  
 RN [4]  
 RP DISULFIDE BONDS.  
 RX MEDLINE=83105095; PubMed=7151781;  
 RA Benkouka F., Guidoni A.A., de Caro J.D., Bonicel J.J.,  
 RA Desnuelle P.A., Roverly M.;  
 RT "Porcine pancreatic lipase. The disulfide bridges and the sulfhydryl  
 RT groups.";  
 RL Eur. J. Biochem. 128:331-341(1982).  
 RN [5]  
 RP SUBSTRATE-BINDING SITE.  
 RX MEDLINE=82000578; PubMed=6791692;  
 RA Guidoni A.A., Benkouka F., de Caro J.D., Roverly M.;  
 RT "Characterization of the serine reacting with diethyl p-nitrophenyl  
 RT phosphate in porcine pancreatic lipase.";  
 RL Biochim. Biophys. Acta 660:148-150(1981).  
 RN [6]  
 RP STRUCTURE OF CARBOHYDRATE.  
 RX MEDLINE=88082841; PubMed=3691527;  
 RA Fournet B., Leroy Y., Montreuil J., Decaro J., Roverly M.,  
 RA van Kuik J.A., Vliegthart J.F.G.;  
 RT "Primary structure of the glycans of porcine pancreatic lipase.";  
 RL Eur. J. Biochem. 170:369-371(1987).  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS), AND REVISIONS TO 30-32.  
 RX MEDLINE=96279347; PubMed=8663362;  
 RA Hermoso J., Pignol D., Kerfelec B., Crenon I., Chapus C.,





Db                    ||| |: |                    | |                    ::| :                    | : |||:| ||  
296 KCFPCPSEGCPQMGHYADRFPGKTNGVVSQVFYLNLTGDASNFARWRYKVSVTLS 348

Search completed: March 4, 2004, 10:25:14  
Job time : 50 secs